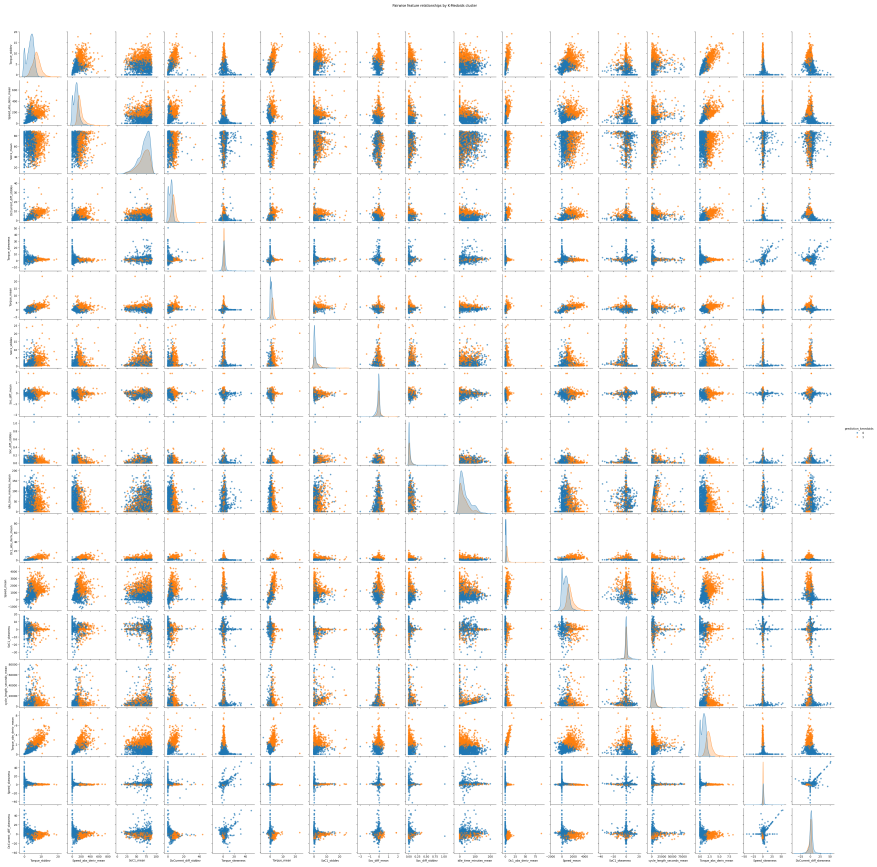
CHALMERS
UNIVERSITY OF TECHNOLOGY



Drive Cycle Analysis for the Electric Powertrain in Terminal Tractors:

A Data-Driven Approach to Performance Evaluation

Master's thesis in Electrical Engineering

Fabian Magnusson
Edwin Marjanovic

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Drive Cycle Analysis for the Electric Powertrain in Terminal Tractors: A Data-Driven Approach to Performance Evaluation

FABIAN MAGNUSSON
EDWIN MARJANOVIC

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2026

Drive Cycle Analysis for the Electric Powertrain in Terminal Tractors: A Data-Driven Approach to Performance Evaluation
FABIAN MAGNUSSON EDWIN MARJANOVIC

Supervisor: Erik Jansson, Volvo Penta
Examiner: Stefan Lundberg, Department of Electrical Engineering

Master's Thesis 2026
Department of Electrical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Pairplot of all features utilized during the feature-based clustering algorithms.

Typeset in LaTeX
Printed by TeknologTryck
Gothenburg, Sweden 2026

Drive Cycle Analysis for the Electric Powertrain in Terminal Tractors: A Data-Driven Approach to Performance Evaluation
FABIAN MAGNUSSON
EDWIN MARJANOVIC
Department of Electrical Engineering
Chalmers University of Technology

Abstract

The transition toward electrified powertrains in industrial vehicles places increased demands on understanding real world operational behavior. Terminal tractors operate under highly variable and transient conditions, making traditional standardized drive cycles insufficient for accurate performance evaluation and optimization. This thesis presents a data driven framework for analyzing, categorizing, and simulating operational drive cycles of electric terminal tractors based on real world field data.

Multivariate time series data collected from electric terminal tractors were preprocessed and segmented into individual drive cycles using application specific operational signals. A set of interpretable features capturing both steady state and dynamic behavior was extracted for each cycle. Dimensionality reduction and feature selection were performed using Principal Component Analysis and Principal Feature Analysis to retain the most informative characteristics while maintaining interpretability. The results were that only 17 principal components out of the original 24 were needed to describe 95% of the explained variance.

Multiple clustering techniques, including Hierarchical Agglomerative Clustering, K-Means, K-Medoids, and a convolutional autoencoder based clustering, were applied and evaluated using internal validation metrics. The resulting clusters revealed two distinct operational regimes, representative usage patterns, and outlying behaviors across the fleet. These two operational regimes were defined as low load and high load, where the low load cluster is defined by its lower variance and more stable values, while the high load cluster is defined by higher variance and a broader range of values in torque for example. Representative and atypical drive cycles from each cluster were subsequently integrated into a simulation model of the electric powertrain to evaluate component behavior, energy usage, and engine efficiency operation under different load configurations.

The results demonstrate that data driven cycle analysis can effectively characterize real world usage patterns and provide valuable insights for how powertrain components, such as the engine and battery, are affected during operation. The proposed methodology offers a scalable framework for leveraging operational data to identify dominant operating regimes and guide structured evaluation of the electrical powertrain in terminal tractors.
Keywords: Drive cycle, unsupervised learning, clustering, K-Means, HAC, K-Medoids, feature extraction, time series signals, powertrain, operational modes, CAE, cluster validation metrics.

Acknowledgements

First of all we would like to express our sincerest gratitude towards everyone that is a part of the Data Science & Engineering department at Volvo Penta who has supported us extensively throughout this thesis. An extra special thank you goes out to our supervisor Erik Jansson for his unending support and valuable insights in making this thesis a reality. Lastly we would like to thank all of the amazing people we met during our time at Chalmers University of Technology, without you we wouldn't have made it.

Edwin Marjanovic & Fabian Magnusson, Gothenburg, April 1, 2026

Contents

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

BEV	Battery Electric Vehicle
BS	Battery System
CAE	Convolutional Autoencoder
EV	Electric Vehicle
EM	Electric Machine
HAC	Hierarchical Agglomerative Clustering
ICE	Internal Combustion Engine
MS	Mechanical System
PCA	Principal Component Analysis
PDU	Power Distribution Unit
PEC	Power Electronics
PFA	Principal Feature Analysis
SAE	The Society of Automotive Engineers
SOC	State of Charge

Nomenclature

Below is the nomenclature of parameters that have been used throughout this thesis.

Parameters

k	Number of clusters
N	Number of samples in a drive cycle
$x_t^{(s)}$	Value of signal s at time index t
C_i	Cluster i
$\|C_i\|$	Number of samples in cluster i
μ	Mean value
$\mu^{(s)}$	Mean value of signal s over one cycle
μ_i	Centroid of cluster i
σ	Standard deviation
skew	Skewness
$\|\Delta\|$	Average absolute discrete derivative
$d(\cdot)$	Distance metric (Euclidean unless otherwise stated)
S_i	Intra cluster scatter for cluster i (Davies Bouldin)
M_{ij}	Distance between centroids of clusters i and j
R_{ij}	Cluster similarity ratio between clusters i and j (Davies Bouldin)
DB	Davies Bouldin index
CH	Calinski Harabasz index

List of Figures

List of Tables

1
Introduction

1.1 Background

With a rapidly changing society that is increasingly focused on data [1], Volvo Penta's mission is to remain at the forefront of the maritime and industrial machinery sectors. Recent acceleration in digitalization enables highly accurate engine data from real world operations. Volvo Penta collects time-series data from a wide variety of applications using data loggers. These loggers extract signals from the powertrain and different controllers from a variety of applications. By understanding the underlying patterns of this data and the needs of their consumers, it creates opportunities for Volvo Penta to further develop and evolve the user experience. This enables Volvo Penta to develop predictive maintenance strategies and highlight opportunities for optimization in powertrain design, performance, reliability, and user experience.

As the industry transitions toward sustainable alternatives, such as electric battery solutions or hydrogen fuels, these analytical capabilities become even more crucial. Today, Volvo Penta has a firm grasp on combustion engines and their powertrains, but is only at the beginning of building equivalent analytical depth for its emerging electrical counterparts. Strengthening this understanding is essential to ensure that future sustainable powertrains maintain the same levels of performance, reliability, and user satisfaction.

In engine and powertrain signals there are patterns that emerge which can be summarized into a drive cycle, that represents how a machine is operated over time. Drive cycles are valuable because they allow engineers to design, optimize, and adapt powertrains to the specific tasks their vehicles perform. They also form the basis for estimating component loads, supporting durability assessments and the development of effective maintenance strategies [2].

With the increasing availability of real-world operational data, recent research shows that machine learning methods, such as clustering and pattern extraction, can automatically derive representative drive cycles from large datasets of user data. These data-driven cycles capture realistic usage patterns more accurately than standardized laboratory profiles and enable more informed decisions about powertrain calibration, energy efficiency, and predictive maintenance.

1.2 Aim

The primary objective of this thesis is to extract, analyze, and compare features and operational patterns utilizing data collected from the powertrain in electrical terminal tractors. By identifying and defining representative and outlying drive cycles from this data, the thesis aims to provide insights for how the powertrain responds to clustered drive cycles. Beyond the extraction and clustering of these patterns, the thesis further aims to link the identified drive cycles to certain powertrain characteristics. Ultimately, the goal is to categorize drive cycles using both real world data and simulation results. This categorization of different operating regimes, together with simulations of the application, enables a deeper understanding of how the application is actually utilized and aims to reveal baseline powertrain behaviors during drive cycle operation.

1.3 Limitations

The scope of the thesis will be limited by a few aspects. First and foremost the data that is to be used originates from field tests, this means that all data has been collected from data loggers on real world applied terminal tractors. The data accumulation will be limited to a time period that ends in September 2025 since the thesis aims to have reproducible results during the work period. Another limitation is that the powertrain for the application is entirely electric, meaning that certain components characteristic for other powertrains, such as hybrid electric vehicles, plug-in hybrids and conventional internal combustion engines, won't be explored.

An accurate Simulink model, representing the terminal tractor, will be provided from Volvo Penta. Due to this, the amount of possible changes regarding values and powertrain components will be limited. Explicit optimization of specific components and/or control strategies was therefore deemed as outside the scope of the thesis. By being provided with a model this also implies that there will not be any validation done on the physical/real world application.

Due to the powertrain architecture including two separate engines, a selective signal analysis strategy was applied. Engine one was chosen as the primary subject of analysis because it features multiple gear ratios and governs gear shift events, making its signals more informative for operational characterization. Engine two operates with a fixed gear ratio and therefore provides less variation in behavior across operating conditions. As a result, a limitation of this study is that detailed signal analysis is largely restricted to engine one. Signals from engine two are only incorporated in comparative metrics, specifically when evaluating inter engine differences in parameters such as state of charge (SOC) and DC current. These comparisons are used solely to assess load sharing and work distribution when both engines operate within a shared operating range, in order to verify whether the load is distributed evenly between them.

Another limitation will be that this thesis only aim to explore the data and powertrain connected to Volvo Penta's terminal tractors and therefore might not be deemed a suitable for all terminal tractors. By excluding in depth GPS-data, see Section 1.4, there will also be limited results and discussion regarding the locations of the terminal tractors/applications. Lastly, due to the thesis analyzing a Volvo Penta owned product, all results may not be eligible for publishing due to secrecy.

1.4 Ethical and environmental aspects

A recurring concern in data collection and handling relates to personal integrity. The logging of geographical data may be perceived as an intrusive practice by application operators, as the tracking and monitoring of employee movements give rise to several ethical issues. However, to enable a comprehensive understanding and optimization of application usage, it is essential to analyze patterns of use and movement. In this thesis, no personal data will be collected or analyzed. Only data necessary for assessing application usage will be utilized.

Another important aspect concerns the environmental impact. The European Union has set the target of achieving climate neutrality with respect to greenhouse gas emissions by 2050 [3]. A central requirement in reaching this objective is the transition from fossil fuels to renewable energy sources, a goal to which Volvo Penta is also committed. This thesis aims to contribute to that effort by defining and analyzing the optimization of Volvo Penta's new electric powertrains for use in electric terminal tractors.

1.5 Specification of the issue being investigated

The thesis aims to investigate how drive cycles can be analyzed, categorized, and used to understand vehicle usage behavior and its implications for powertrain design and optimization. The main research questions guiding this work are:

- **How can different types of drive cycles be extracted, and which signals and features are most relevant to accurately describe usage behavior?**
- **What groups of cycles can be identified through clustering, and how can outliers or atypical behaviors be characterized?**
- **Which clustering algorithms are suitable for analyzing and segmenting the drive cycles and in what way do they yield meaningful insights?**
- **Is it possible to incorporate the extracted drive cycles into a simulated model of the application?**
- **How can the identified drive cycle categories and their simulations inform powertrain evaluation and future optimization efforts?**

1.6 Related Work

This thesis work is based on a previous thesis, conducted by Rasmus Hellrand and Jakob Malmer Göransson, at Volvo Penta during the spring of 2025 [4]. Their approach focused on utilizing unsupervised learning and statistical techniques to define and segment drive cycles. Their aim was to use clustering methods to differentiate between different types of engine cycles across one specific engine applications.

To expand and improve on their previous work, this thesis utilizes two electrical engines from six different applications whilst theirs only used one diesel engine from one application. Our thesis also explores the use of simulations as a complement to clustering analysis to further broaden the analysis. Furthermore, the availability of their implemented methodology presents significant value for feature extraction and clustering, while also offering scope for further refinement.

2
Theory

In the following sections of this chapter, there will be in depth explanation of the used theory for the conducted thesis. The theory will cover; *Time series, Electric powertrains, Feature engineering, Clustering methods, Clustering Evaluation and Simulations.*

2.1 Time series

A time series is a sequence of observations, usually ordered in time, but in some cases the ordering may be in regard to other dimensions. The dominant aspect of time series, the one that distinguishes it from other statistical analyses, is the importance of order in which the observations are made/ take place [5]. There are two types of time series, univariate and multivariate.

A univariate time series is a time series consisting of observations that has been sequentially recorded over the same time increment for a certain interval [6]. The univariate time series is represented as:

$$\boldsymbol{X} = [X]^K \tag{2.1}$$

where K equals the number of observed number of time instances.

A multivariate time series is defined as a set of, at least two, independent observations which are univariate time series. The observed variables may be related and may exhibit both dependencies and interactions over time [7]. A multivariate time series, with T number of timestamps, is represented as:

$$\boldsymbol{X} = [X_1, X_2, ..., X_T] \tag{2.2}$$

where:

$$X_t \in \Re^n \tag{2.3}$$

is observed at timestamp t with n number of dimensions, where $n>0$ [7].

2.2 Electric powertrains

An electric vehicle (EV) refers to a vehicle that relies on electric power for motion. This power is most commonly delivered by either a battery or a fuel cell. An EV powertrain refers to the system within the vehicle that converts the electrical energy into mechanical motion [8]. The following components are all part of an electric powertrain [8]:

- Energy storage - battery or fuel cell
- Power electronics (PEC) - inverter/converter for controls
- Electric machine (EM) - converts energy (electric motor)
- Mechanical system (MS) - transmission system.
- Drive shaft (and its mechanical parts)

When comparing an EV with a conventional internal combustion engine (ICE) the EV has less environmental impact due to a negligible tailpipe emissions as well as a higher energy efficiency [8]. Battery electric vehicles (BEVs) are often referred to as *pure-electric vehicles*, and all their power originates from the stored energy in the rechargeable battery [8]. A BEV doesn't use any kind of secondary storage sources, such as, i.e. a fuel tank, and in order to charge the battery, an external energy source must be connected [8].

2.2.1 Powertrain design/lay out

As the core system in a BEV, the powertrain system is similar to the engine and transmission system in a traditional diesel or petrol-fueled vehicle [9]. Nevertheless, BEV are different from traditional ICE when it comes to energy conversion [9]. To facilitate understanding, Figure 2.1 shows a diagram of the power transmission process in a battery driven terminal tractor. This is the terminal tractor used during the study and the figure is published with courtesy of Volvo Penta [10]. The Figure 2.1 highlights the BEV powertrain system and the main component being the battery system, power distribution unit (PDU), the power electronics and a motor drive. Figure 2.1 contains: **1** - Active Cooling Unit, **2** - Battery Pack 1 and 2, **3** - DC/DC converter, **4** - Electric Motors 1 and 2, **5** - Onboard bi-directional charger and **6** - Transmission.

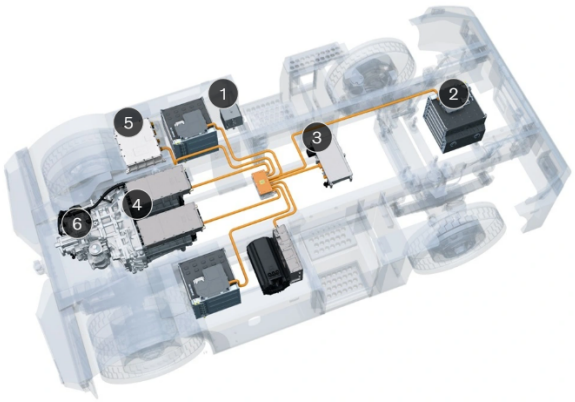


Figure 2.1: A schematic overview of the terminal tractor used during this thesis, with the different components marked by numbers. The following components are: 1 - Active Cooling Unit, 2 - Battery Pack 1 and 2, 3 - DC/DC converter, 4 - Electric Motors one and two, 5 - Onboard bi-directional charger and 6 - Transmission. The figure is published with courtesy of Volvo Penta [10].

Under normal work conditions, for the BEV, the electric energy, stored in the battery system (BS) is firstly delivered to the PDU [9]. From the PDU the energy is distributed further, in to the motor controller and lastly transformed into mechanical energy used to operate the BEV [9].

2.2.2 Torque-speed characteristics for an electric vehicle.

The torque speed characteristics, for an electric vehicle traction motor, describes how the available torque varies with the rotational speed [11]. A key component for the determination of vehicle performance and efficiency [11]. Electrical machines are capable, unlike ICEs, to produce a high torque directly from start, making them ideal for applications where frequent starts-stop periods occur under high load [11].

At low rotational speeds, the electric machine operates in a constant torque region [11]. In this region the torque is mainly limited by the maximum allowed current through the motor and power electronics [11]. The low speed operating region enables strong launch and high traction at low vehicle speeds [11]. After the initial low speed region, the electric machine typically transitions into a region of constant power where the torque decreases with approximately the inverse of the speed in order to keep the output power constant [11]. The constant power region enables the vehicle to operate at higher speeds without requiring additional mechanical gearing [11]. The torque speed characteristic directly affects the operating points for the electric machine during operation and when combined with efficiency maps, gives insight in how often the machine operates in regions of high- and low efficiencies [11]. For a vehicle applications such as a electric terminal tractor, characterized by transient- and repeated acceleration events, understanding the torque speed relationship is essential [11].

2.2.3 Electric machine efficiency map interpretation.

Electric vehicle traction motors operate over a wide range of torque and speed, making efficiency characterization essential for electric powertrain analysis [12]. Efficiency maps are frequently used to describe motor performance by relating torque and rotational speed to corresponding efficiencies across the operation. This allows for a possibility of energy consumption evaluation under realistic load conditions [12]. The efficiency maps are typically generated using steady state analytical- or finite element based models enabling efficient computation, but at the cost of transient effects during rapid changes in torque and speed being neglected [12] [13].

In order to address the neglected effects, time stepping- and dynamic simulation approaches evaluate the operating efficiency during actual operating drive cycles. These methods enable the capturing of events reflecting acceleration and deceleration [13]. By combining torque-speed characteristics with efficiency mapping and transient analysis a more accurate representation of the electric machine operating points are detected, resulting in an improved assessment regarding energy consumption and powertrain performance [12] [13].

2.2.4 Battery discharge during transient conditions.

Battery discharge during transient conditions indicates operating events where the current changes rapidly, these events could be accelerating phases or fast shifts between operating modes [14]. During the transient phases, the voltage response is strongly influenced by short-lived internal dynamics such as polarization and relaxing processes [14]. This results in an instantaneous voltage drop reflecting both the electrochemical state as well as the dynamic resistive effects [14].

Transient discharge behavior is closely related to thermal dynamics since rapid changes in current directly affects the generated heat within the battery [15]. During realistic drive cycle operation, the battery temperature should progress dynamically due to the combined effects of heat generation, operation of the cooling system and the surrounding vehicle temperature [15]. Models capturing transient thermal events, taking conduction, convection, radiation etc. into consideration has shown to accurately predict battery surface temperature under real vehicle operation. These models can be used to assess extended thermal and degradation risks for the battery [15]. The models imply that transient discharges should be treated as an electro thermal problem where the electrical- and thermal states develop together influences each other [15].

2.3 Feature engineering and clustering

The Society of Automotive Engineers (SAE) is an organisation that offers several papers, standards, e-books etc. regarding automotive and aerospace engineering as well as helping with producing these standards [16]. The produced standards are key components in defining several descriptive drive cycle features where most features are related to velocity and acceleration as well as the number of start and stops and the duration of idle periods. A drive cycle and its corresponding features is a representation of the usage for a vehicle where the drive pattern, road characteristics and traffic characteristics are taken into consideration [17]. Drive cycles are typically used to assess different performance parameters and system sizing for vehicles and have been an essential part of emission testing and simulation tests for decades [17].

To utilize the features from the drive cycles, feature engineering can be applied. Feature engineering uses different techniques typically applied after collecting and cleaning the input data [18]. During the cleaning process one typically deals with missing values, errors, outliers etc. After the cleaning, feature engineering is used to process the data, from raw measurements into descriptive features [18]. The use of feature engineering aims towards the development of smart features; either by modification of already existing features or by creating new features such as calculating the average time of a series segment [18].

To further make use of these features, clustering can be applied to evaluate how, and if, it is possible to separate the data and drive cycles into different interpretable groups. Clustering algorithms group data instances into a number of clusters, where the items within each cluster should be similar to one another while being distinctly different from those in other clusters. These similarities and dissimilarities must be defined in a clear and meaningful way [19]. Because clustering relies on grouping similar points together, it is necessary to evaluate how appropriate these groupings are. This evaluation is typically based on two main types of measures: distance measures and similarity measures [20]. The subsections following will cover more about how feature engineering and clustering algorithms work and what defines them.

2.3.1 Principal Component Analysis

Principal Component Analysis (PCA) is an exploratory method for dimensionality reduction that aims to represent a dataset with fewer variables while preserving as much variation as possible [21]. For a dataset with n observations and p numerical variables arranged in a data matrix, PCA constructs new variables, called principal components, as linear combinations of the original variables.

The first principal component is defined as the direction that maximizes the variance of the projected data under a unit length constraint. This leads to an eigenvalue problem of the sample covariance matrix, where the eigenvectors define the component loadings and the corresponding eigenvalues quantify the variance explained by

each component. Subsequent components are computed in decreasing order of explained variance and are constrained to be orthogonal to earlier components, which makes them uncorrelated.

In practice, PCA is typically applied to mean centered data and can be computed using singular value decomposition of the centered data matrix. Retaining only the first q components yields a lower dimensional representation that is optimal in a least squares sense. The proportion of total variance explained by the retained components is commonly used to determine how many components to keep and to assess the quality of the reduced representation.

2.3.2 Principal Feature Analysis

Principal Feature Analysis (PFA) is a feature selection technique that builds on PCA to identify a subset of representative original features while reducing redundancy [22]. Instead of constructing new transformed variables, PFA selects a smaller subset of existing features that captures the dominant variation structure of the dataset. n The method begins by applying PCA to a standardized feature matrix. The resulting principal components define orthogonal directions of decreasing variance. Each original feature can be represented by its loadings across the retained principal components, which describe how strongly the feature contributes to each major mode of variation.

To give greater importance to components that explain more variance, the loading vectors are weighted by the square root of the explained variance of each retained component. The weighted loading vectors can also be standardized to ensure comparable scale across dimensions. In this transformed principal component space, each original feature is represented as a point defined by its weighted loadings.

Clustering is then performed on these feature points, commonly using K means clustering. The number of clusters is typically chosen equal to the number of retained principal components, reflecting the interpretation that each component represents a dominant variation pattern. From each cluster, the feature closest to the cluster centroid is selected as the representative feature. The resulting subset provides a compact set of features that preserves the main structural information in the data while reducing redundancy.

2.3.3 K-Means

K-Means is a partition based unsupervised learning method and is often referred to as one of the most popular data mining algorithms, which may explain its extensive use within the research community [23]. The objective of K-Means is to divide a dataset into a predetermined number of clusters k, such that observations within the same cluster are more similar to each other than to observations in different clusters, according to a chosen distance measure.

In this work, the Euclidean distance is used, and it is calculated as:

$$\|x - c_i\| = \sqrt{\sum_{j=1}^{n}(x_j - c_{ij})^2} \tag{2.4}$$

where $\|x - c_i\|$ denotes the distance between a data point x and the cluster center c_i in Euclidean space and n is the number of features (dimensions) in the data [24]. The pseudocode used throughout this thesis follows the conventions established in Cormen et al. [25], where algorithms are expressed using structured English-like constructs, independent of any specific programming language.

Algorithm 1 Pseudocode for clustering using K-Means

procedure K-MEANS(X, k) ▷ X is the set of data points and k is the number of clusters
  Description of input: dataset $X = \{x_1, x_2, \ldots, x_n\}$ where each $x_i \in \mathbb{R}^n$, and k is the predetermined number of clusters.
  Description of output: a partition of X into k clusters $\{C_1, C_2, \ldots, C_k\}$ and the corresponding cluster centers $\{c_1, c_2, \ldots, c_k\}$.
  Initialize cluster centers $c_1, \ldots, c_k$ (e.g., randomly selected points from X).
  repeat
    for each point $x \in X$ **do**
      Assign x to the closest center:

$$\text{assign } x \text{ to } C_i \text{ where } i = \arg\min_{j \in \{1,\ldots,k\}} \|x - c_j\|$$

      Where, $\arg\min_{j\in\{1,\ldots,k\}}$ denotes the index j that minimizes the expression [25].
    end for
    for each cluster $i \in \{1, \ldots, k\}$ **do**
      Update center as the centroid of its assigned points:

$$c_i \leftarrow \frac{1}{|C_i|} \sum_{x \in C_i} x$$

      where $|C_i|$ denotes the number of points in cluster C_i.
    end for
  until cluster assignments (or centers) do not change.
  return $\{C_1, C_2, \ldots, C_k\}$ and $\{c_1, c_2, \ldots, c_k\}$.
end procedure

Despite the wide use of K-Means the algorithm still faces some limitations. K-Means struggles with clusters that have random initialization of the centroids which leads to unexpected convergence [23]. Furthermore methods such as K-Means, require the number of clusters, k, to be predefined which is responsible for the cluster shapes and effects from the outliers. Another fundamental problem of the K-Means algorithm is its inability to handle various data types [23]. Additionally, while K-Means is a

clustering method used for minimizing the variance within the clusters, the random selection of cluster centers can often lead to poor clustering results [23].

The K-Means++ algorithm modifies the initialization stage of the conventional K-Means algorithm to yield improved theoretical and empirical performance. It is intended to be used as a seeding method preceding a standard K-Means iteration [26]. K-Means++ distributes the initial cluster centers based on each data point's contribution to the overall inertia, where inertia is defined as the within-cluster sum of squared Euclidean distances, corresponding to the cost function minimized by K-Means. The first cluster center in K-Means++ is selected uniformly at random from the dataset, meaning that each data point has an equal probability of being chosen. Subsequent centers are selected sequentially using the Euclidean distance. For convenience, the distance from a data point x to its nearest previously selected center is defined as:

$$D(x) = \min_j \|x - c_j\|. \tag{2.5}$$

The notation $c_i^{(j)}$, see algorithm 2, denotes the j-th candidate cluster center sampled for the i-th center during the initialization procedure. The probability of selecting a data point as a candidate center is proportional to the squared distance $D(x)^2$, such that points located farther from existing centers are more likely to be selected. Squaring the distance emphasizes separation between centers and is consistent with the squared-distance formulation of the K-Means objective function. The probabilities are normalized to ensure that they sum to one over the dataset. As a result, K-Means++ produces a well-scattered set of initial cluster centers across the data space. The K-Means++ algorithm is explained in Algorithm 2.

Algorithm 2 Pseudocode for clustering using K-Means++

procedure K-MEANS++(X, k, t) ▷ X is the set of data points and k is the number of clusters
  Description of input: Dataset $X = \{x_1, x_2, \ldots, x_n\}$, predetermined number of clusters k, and number of trials t.
  Description of output: Initial cluster centers $\{c_1, c_2, \ldots, c_k\}$.
  Select the first center c_1 uniformly at random from X. ▷ Each data point has equal probability of being selected.
  for $i = 2$ to k **do**
    for each data point $x \in X$ **do**
      Compute the distance to the nearest selected center:

$$D(x) = \min_{j \in \{1,\ldots,i-1\}} \|x - c_j\|$$

    end for
    Initialize an empty candidate set S.
    for $j = 1$ to t **do** ▷ Perform t sampling trials
      Sample a candidate center $c_i^{(j)} \in X$ with probability

$$P(x') = \frac{D(x')^2}{\sum_{x \in X} D(x)^2}$$

      where $x' \in X$ denotes a candidate data point sampled from the dataset.
      Add $c_i^{(j)}$ to S.
    end for
    Select $c_i \in S$ that minimizes the sum of squared distances to the nearest existing center.
  end for
  Use the selected centers to initialize the standard K-Means algorithm.
  return Initial cluster centers $\{c_1, c_2, \ldots, c_k\}$.
end procedure

2.3.4 K-Medoids

The K-Medoids clustering algorithm is a partition based unsupervised learning method closely related to K-Means [27]. The objective is to divide a dataset into a predetermined number of clusters based on similarity or distance measures. Unlike K-Means, which represents each cluster by the mean of its assigned points, K-Medoids selects an actual data point as the cluster center, referred to as the medoid. This makes the method more robust to noise and outliers, since medoids are always valid observations from the dataset [27]. The K-Medoids algorithm operates by iteratively assigning data points to the nearest medoid and updating medoids based on a cost minimization criterion. A general outline of the algorithm is given in Algorithm 3.

Algorithm 3 Pseudocode for clustering using K-Medoids

procedure K-MEDOIDS(X, k) ▷ X is the set of data points and k is the number of clusters
  Description of input: dataset $X = \{x_1, x_2, \ldots, x_n\}$ where each $x_i \in \mathbb{R}^n$, and k is the predetermined number of clusters.
  Description of output: a partition of X into k clusters $\{C_1, C_2, \ldots, C_k\}$ and the corresponding medoids $\{m_1, m_2, \ldots, m_k\}$, where each $m_i \in X$.
  Initialize medoids $m_1, \ldots, m_k$ by selecting k points from X (e.g., uniformly at random).
  repeat
    for each point $x \in X$ **do**
      Assign x to the nearest medoid:

$$\text{assign } x \text{ to } C_i \text{ where } i = \arg\min_{j \in \{1,\ldots,k\}} \|x - m_j\|$$

    end for
    for each cluster $i \in \{1, \ldots, k\}$ **do**
      Update medoid by selecting the point in C_i that minimizes total distance:

$$m_i \leftarrow \arg\min_{m \in C_i} \sum_{x \in C_i} \|x - m\|$$

    end for
  until medoids $\{m_1, \ldots, m_k\}$ do not change.
  return $\{C_1, C_2, \ldots, C_k\}$ and $\{m_1, m_2, \ldots, m_k\}$.
end procedure

2.3.5 Hierarchical clustering

As mentioned in the section, K-Means is a common and well utilized method for clustering but it has it's shortcomings. One approach to address these limitations is hierarchical clustering, which organises data into a tree-like structure. Each node in the tree represents a single data point, and the root contains the entire dataset. Nodes in between represent intermediate groupings of varying sizes. The core principle is to nest clusters within larger clusters, building the hierarchy either from the bottom up or the top down. There are two main conceptual approaches to hierarchical clustering. Hierarchical Agglomerative Clustering (HAC) takes a bottom-up approach, beginning with individual data points and progressively merging them. Divisive clustering, on the other hand, takes a top-down approach, starting with one all-encompassing cluster and repeatedly splitting it [28].

HAC begins by treating every individual data point as its own cluster. At each step, the two most similar clusters are identified and combined into a single new cluster, replacing the two that were merged. This procedure repeats until all data points have been merged into one cluster, or some stopping condition is reached [28][29]. The result can be represented as a dendrogram, a tree diagram that visualizes at what point of the algorithm that the clusters were merged and how they relate, the

steps for HAC is presented in Algorithm 4.

Algorithm 4 Pseudocode for Hierarchical Agglomerative Clustering

procedure HIERARCHICAL AGGLOMERATIVE CLUSTERING(X, d) ▷ X is the dataset and $d(\cdot, \cdot)$ is a dissimilarity measure
  Input: dataset $X = \{x_1, x_2, \ldots, x_n\}$ where each $x_i \in \mathbb{R}^n$, and a pairwise dissimilarity function $d(x_i, x_j)$.
  Output: a hierarchical clustering represented as a dendrogram (linkage matrix) L.
  Initialize each data point as a singleton cluster:

$$\mathcal{C} \leftarrow \{\{x_1\}, \{x_2\}, \ldots, \{x_n\}\}$$

  Initialize an empty linkage list L.
  while $|\mathcal{C}| > 1$ **do**
    Find the two closest clusters:

$$(C_a, C_b) \leftarrow \arg\min_{C_i \neq C_j} d(C_i, C_j)$$

    Record merge information in L:

$$L \leftarrow L \cup (C_a, C_b, d(C_a, C_b))$$

    Merge clusters:

$$C_n \leftarrow C_a \cup C_b$$

    Remove C_a and C_b from $\mathcal{C}$ and add C_n.
    Update distances between C_n and all remaining clusters according to the chosen linkage criterion.
  end while
  return linkage matrix L.
end procedure

One of the advantages of using hierarchical clustering is that there is no requirement to specify or determine the number of clusters in advance, unlike K-Means [28]. Thanks to this characteristic, hierarchical clustering is beneficial when it comes to drive cycle analysis. Due to the unknown behaviors of the drive cycles, approaching the analysis with hierarchical clustering allows for an exploratory approach in order to group cycles with similar characteristics together. These groups could, for example, be low-load drive cycles, normal-load drive cycles, and high-load drive cycles.

To calculate and determine the distance between the clusters that are being merged during HAC, it is possible to make use of linkage methods [30]. In Algorithm 4, the linkage criterion is utilized when identifying the two closest clusters to merge, and when updating the distances between the newly formed cluster and the remaining clusters. There are different methods that can be used in order to calculate linkage, common ones being: single linkage, average linkage, complete linkage and Ward

linkage. The distance $d(C_i, C_j)$ between clusters C_i and C_j is computed differently for each method as follows [30]:

- Single linkage – the distance is the minimum distance between any two points across the clusters:
$$d(C_i, C_j) = \min_{x \in C_i,\, y \in C_j} \|x - y\| \tag{2.6}$$
- Average linkage – the distance is the average distance between all pairs of points across the clusters:
$$d(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{x \in C_i} \sum_{y \in C_j} \|x - y\| \tag{2.7}$$
- Complete linkage – the distance is the maximum distance between any two points across the clusters:
$$d(C_i, C_j) = \max_{x \in C_i,\, y \in C_j} \|x - y\| \tag{2.8}$$
- Ward's linkage – the distance is the increase in within-cluster variance after merging, computed as:
$$d(C_i, C_j) = \sqrt{\frac{2|C_i||C_j|}{|C_i| + |C_j|}} \|\mu_i - \mu_j\| \tag{2.9}$$

where μ_i and μ_j are the centroids of clusters C_i and C_j respectively.

Each linkage method has different properties, making them suitable for different types of data. Single linkage clustering is suitable for tracing irregular, non-elliptical shapes. This, however, comes at the cost of being less suitable for noisy data containing outliers [31]. Average linkage offers a balance between single and complete linkage, making it less sensitive to outliers than single linkage while still capturing broader cluster structure. However, it can struggle when clusters differ significantly in size or density [31]. Contrary to the single linkage method, complete linkage is more robust and better suited for noisy data with outliers. This, however, comes at the cost of potentially biased results when applied to large and spherical clusters [31]. As a result, complete linkage often produces more compact clusters than single linkage. Ward's linkage is most commonly used for quantitative data due to its resistance to noise and outliers [31].

2.3.6 Autoencoders

Autoencoders are a specific type of neural network that is designed to reconstruct the input at the output [32]. This is done through encoding and decoding at a hidden layer that is referred to as the bottleneck. This bottleneck is the crucial part of an autoencoder since it forces the network to learn and prioritize the most representative aspects of the input by ignoring noise and irrelevant details. The encoder maps the input to a latent feature space which usually is lower in dimension than the output. The decoder then tries to reconstruct the input from the compressed data.

The value in using autoencoders as opposed to conventional clustering algorithms lies in its ability to utilize the raw time series as input, which preserves the temporal structure that is otherwise lost when reducing each cycle into statistical features.

A Convolutional autoencoder (CAE) is a special autoencoder that utilizes the robust structure of a convolutional neural network by applying convolutional and spatial layers to reduce dimensionality and compress the input into a feature rich representation [33]. A general structure of a CAE can be seen in Figure 2.2. The CAE is built from two cooperating parts, an encoder and a decoder, each using convolution based layers. The encoder processes the input, and gradually transforms it into progressively more abstract feature representations through stacked one dimensional or multi dimensional convolutions, together with pooling operations and nonlinear activations, for example, a Rectified Linear Unit (ReLU). ReLU activations are commonly used because it helps stabilize gradient flow during training and encourages sparse feature representations. Pooling stages reduce the resolution of feature maps, and the most compressed internal representation, the bottleneck (latent space) appears at the deepest encoder layer, where the core informative structure of the input is captured. The decoder then performs these steps in reverse, reconstructing the input from the compact latent representation. It expands the feature maps step by step using transposed convolution layers, which increases the resolution and restores the spatial structure. Because convolutional operations maintain local spatial or temporal relationships, convolutional autoencoders can learn hierarchical features that become more abstract with network depth.

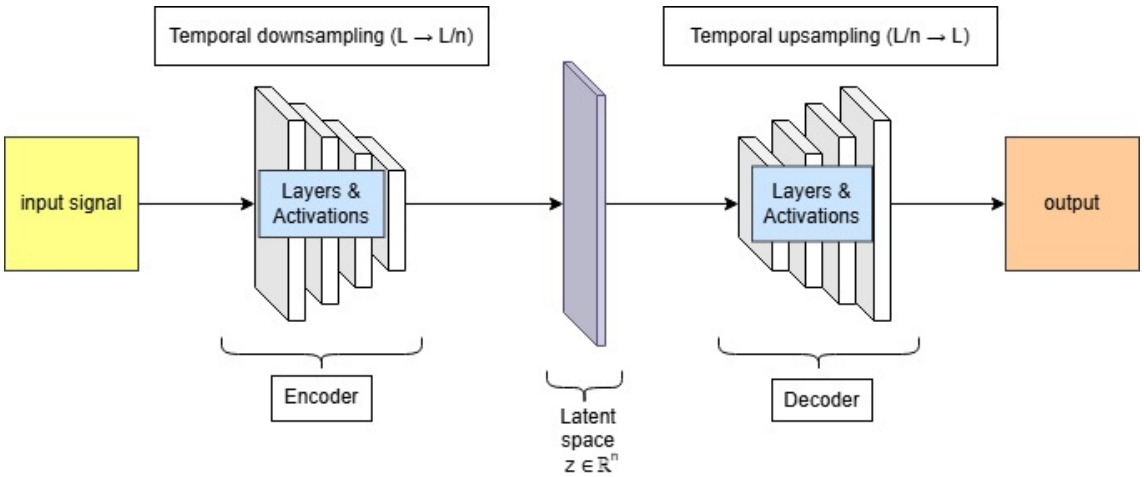


Figure 2.2: General structure of a CAE, where n is the temporal downsampling and upsampling factor applied to the signal length.

2.3.7 Evaluating Clusters

To begin with, clustering is typically divided into two categories, supervised clustering and unsupervised clustering. Due to the unsupervised nature of clustering, comparing and evaluating methods are a difficult task since usual clustering challenges (e.g., no labels) apply [34]. Following this, evaluation often relies on internal

metrics (compactness, separation) or external when labels are available. However, time-series clustering has additional complexities (i.e., temporal distortions, variable lengths) which make evaluation non-trivial. Evaluation is still possible but one has to be conscious about the choice of similarity measure since it directly affects the cluster outcome. The evaluation must consider whether the representation highlights meaningful structure (i.e., representation quality and cluster quality) [34]. Some common evaluation metrics are Silhouette Score, Calinski-Harabasz index and the Davies-Bouldin Index, these are listed below.

Silhouette Score

Silhouette score, is a value, between -1 and 1, assigned to each data point in a cluster. This score is based on the tightness and separation of the data points within the clustering. Values closer to $+1$ indicates that a data point is well matched to its own cluster, values close to 0 suggest that the data point lies near the boundary between clusters, and negative values indicate that the data point may be assigned to the wrong cluster. The silhouette score measures two key aspects of clustering: cohesion, which quantifies how close data points within a cluster are to each other, and separation, which assesses how distinct clusters are from one another. The Silhouette score is composed of two different scores:

- a_i - The mean Euclidean distance between data point i and all other data points within the same cluster C_k, and it is calculated as:

$$a_i = \frac{1}{|C_k| - 1} \sum_{j \in C_k, j \neq i} \|x_i - x_j\| \tag{2.10}$$

- b_i - The mean Euclidean distance between data point i and all data points within the next nearest cluster C_m, and it is calculated as:

$$b_i = \min_{m \neq k} \frac{1}{|C_m|} \sum_{j \in C_m} \|x_i - x_j\| \tag{2.11}$$

For each data point i, the silhouette value s_i is defined as

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \tag{2.12}$$

The silhouette score for a cluster C_k with n_k data points is computed as the mean silhouette value of its data points:

$$S_k = \frac{1}{n_k} \sum_{i \in C_k} s_i \tag{2.13}$$

The overall silhouette score for the clustering is given by

$$S = \frac{1}{N} \sum_{i=1}^{N} s_i \tag{2.14}$$

Where N is the total amount of data points.

Calinski-Harabasz index

If there are no known ground-truth labels, one can use the Calinski-Harabasz (CH) index, also known as the Variance Ratio Criterion [35]. The CH index measures the quality of the clustering by comparing the total within cluster dispersion across all clusters with the dispersion between cluster centroids. The within cluster dispersion is calculated as:

$$W_k = \sum_{i=1}^{k} \sum_{x \in C_i} (x - c_i)(x - c_i)^T \tag{2.15}$$

where W_k is the total within-cluster dispersion, C_i is the set of points in cluster i with the centroid c_i [35]. The dispersion between the cluster can be calculated as:

$$B_k = \sum_{i=1}^{k} n_i (c_i - c_E)(c_i - c_E)^T \tag{2.16}$$

where B_k is the total between-cluster dispersion, n_q is the number of points in cluster i with centroid c_i, and c_E is the centroid of the global dataset E, computed as:

$$c_E = \frac{1}{n_E} \sum_{x \in E} x \tag{2.17}$$

where n_E is the total number of data points [35]. Finally, combining W_k and B_k one can calculate the CH-index as:

$$CH = \frac{tr(B_k)}{tr(W_k)} \times \frac{n_E - k}{k - 1} \tag{2.18}$$

where $tr(B_k)$ and $tr(W_k)$ denote the traces of the total between cluster and total within cluster dispersion matrices, and k is the number of clusters [35]. The trace of a matrix, denoted as $\mathrm{tr}(\cdot)$, is defined as the sum of its diagonal elements. For an $m \times m$ matrix B it is calculated as:

$$\mathrm{tr}(B) = \sum_{i=1}^{m} b_{ii} = b_{11} + b_{22} + \cdots + b_{mm} \tag{2.19}$$

The CH index has no fixed upper bound and is interpreted comparatively. Higher values of the CH index indicate a clustering structure with compact clusters and well separated centroids, corresponding to a higher quality clustering. Conversely, lower CH values suggest poor cluster separation and higher within cluster dispersion. In practice, the CH index is commonly used to compare clustering solutions obtained with different numbers of clusters k, where the value of k that maximizes the CH index is considered to provide the most appropriate clustering structure [36].

Davies-Bouldin Index

The Davies-Bouldin index (DB) is a cluster validation metric used to evaluate clustering quality based on intra-cluster compactness and inter-cluster separation, where lower values indicate better clustering performance [37]. DB is defined in terms of within cluster dispersion and inter-cluster separation. DB uses a lower bound of zero and no fixed upper bound. Lower DB values indicate compact clusters and dissimilarity between clusters. There is no threshold for what constitutes a good DB value, much like the CH index. It is primarily used to compare clustering solutions with different numbers of clusters k. In practice, the value of k that minimizes the DB index is selected as the most appropriate clustering configuration [38]. For each cluster i, the within-cluster dispersion S_i is defined as the average distance between all samples in a cluster and the corresponding cluster centroid:

$$S_i = \frac{1}{|C_i|} \sum_{x \in C_i} ||x - c_i|| \tag{2.20}$$

where $|C_i|$ denotes the number of points in cluster i and c_i its centroid. The separation between two clusters i and j is defined as the euclidean distance between their centroids:

$$M_{ij} = ||c_i - c_j|| \tag{2.21}$$

Using these quantities, the pairwise cluster similarity measure R_{ij} is computed as:

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \tag{2.22}$$

This ratio increases when clusters are more scattered internally or when their centroids are closer together, both of which indicate weaker cluster structure. For each cluster i, the maximum similarity with respect to all other clusters is selected:

$$R_i = \max_{j \neq i} R_{ij} \tag{2.23}$$

The Davies–Bouldin index for a dataset with k clusters is then defined as:

$$DB = \frac{1}{k} \sum_{i=1}^{k} R_i \tag{2.24}$$

A limitation of the DB index is that it depends on the chosen distance metric and centroid representation, which can affect comparability across different feature scalings and clustering methods.

3
Methodology

The methodology shown in Figure 3.1 was used to extract, define, and analyze clusters, as well as to simulate the powertrain and it's behavior using data gathered from Volvo Penta's electrical terminal tractors. The first step of the method was to process the raw time series data extracted from the applications, then define what a drive cycle is and begin extracting features from the time-series data. The data was processed using techniques such as similarity measures and feature extraction. As the project progressed, more sophisticated techniques were employed and the most successful methods served as the cornerstone of the thesis.

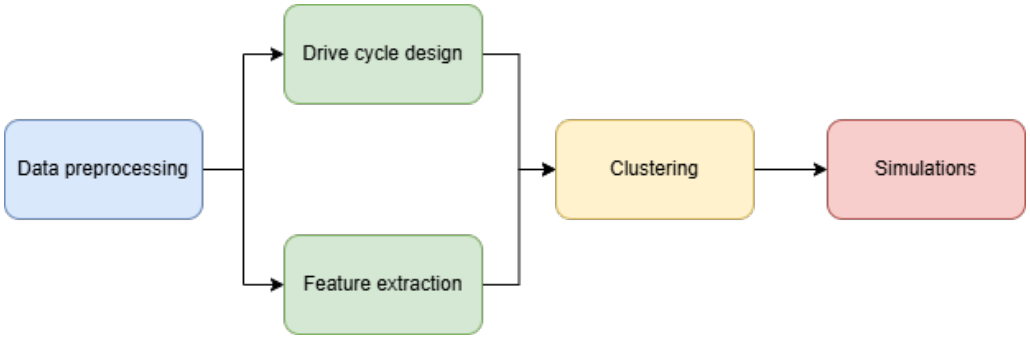


Figure 3.1: Flowchart describing the overview of the methodology.

3.1 Data preprocessing

The first step of the project was to preprocess the data. The initial data is stored in Delta tables, a set of tables good at handling large amount of data, in Volvo Penta's Microsoft cloud server Azure. Each Delta table includes multivariate time-series data, from six different terminal tractors. Signals such as speed, torque, current, etc, that were deemed interesting was extracted. By talking to domain experts at Volvo Penta, reasonable ranges of signal values were defined and the signals was cleaned accordingly by removing values outside of this range. During preprocessing, the applications were assigned random aliases to protect sensitive information. The applications are hereafter referred to as MightyBear_1, CrazyOtter_3, BraveHawk_4, CharmingFox_5, ElegantPanther_6, and CleverWolf_7. Application 2 was excluded from further analysis due to a low number of extracted signals and substantial differences in signal availability compared to the remaining applications.

3.2 Drive cycle design

After the signals were filtered, and the baseline delta table was created, the drive cycle had to be defined. The drive cycles were defined by utilizing a signal that defines different engine modes, the mode utilized for the drive cycle design defines electric activity in the vehicle. This filtering step was made in agreement with Volvo Penta domain experts. Therefore a switch in this signal was defined as the start or end of a cycle. After this the cycle length and time between cycles was calculated. To make sure the cycles were meaningful they were then filtered on the length of the cycle. The thresholds for these filtering steps was set to 15 minutes for the lower boundary and 24 hours for the upper boundary. A cycle shorter or longer than these boundaries were considered irrelevant for further analysis due to the short and/or implausible nature of such driving patterns. After this filtering step, idle features such as idle count and idle time was calculated. An idle period was defined as a part of a cycle where the rotational speed (RPM) was above -100 and below 100 for a time period of more than five minutes and less than 120 minutes. The number of times this occurred was flagged and counted.

3.3 Feature extraction

A signal is a measured time series variable recorded during a drive cycle, while a feature is a numerical summary derived from that signal over the full cycle. In this part of the method, statistical and dynamic features are computed from selected signals to obtain a compact per cycle representation suitable for clustering. The selected signals are shown and described in Table 3.1. Feature selection was performed in collaboration with Volvo Penta domain experts to balance feature complexity and interpretability. Although more rich complex features are possible, a trade off was made to support reliable clustering performance while still maintaining interpretability for real world usage. For each drive cycle, a fixed set of statistical features was computed for each included signal, resulting in one aggregated feature vector per cycle.

Table 3.1: Chosen signals to be used.

Signal	Description
Dc1	Input current to engine 1
Dc2	Input current to engine 2
Speed	Rotational speed for engine 1
Torque	Output torque from engine 1
SOC1	State of charge for battery 1
SOC2	State of charge for battery 2
cycle_length_seconds	Length of a drive cycle in seconds
Dc_diff	Difference between Dc1 and Dc2 at a specific timestamp
idle_time_minutes	Number of minutes the drive cycle is idle
idle_count	Number of idle periods that begin within a drive cycle
SOC_diff	Difference in SOC between battery 1 and battery 2 at a specific timestamp

Let a given drive cycle contain N sampled time points and let $x_t^{(s)}$ denote the sampled value at time index t for signal s. All statistical measures are computed independently for each signal and each cycle. The mean value for signal s in a cycle is defined as:

$$\mu^{(s)} = \frac{1}{N} \sum_{t=1}^{N} x_t^{(s)} \tag{3.1}$$

The standard deviation (σ) is computed as:

$$\sigma^{(s)} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left(x_t^{(s)} - \mu^{(s)} \right)^2} \tag{3.2}$$

The skewness (Skew), which measures the asymmetry of the value distribution within the cycle, is defined as:

$$\text{skew}^{(s)} = \frac{1}{N} \sum_{t=1}^{N} \left(\frac{x_t^{(s)} - \mu^{(s)}}{\sigma^{(s)}} \right)^3 \tag{3.3}$$

To capture dynamic behavior, the magnitude of the discrete derivative ($|\Delta|$) was also computed. Using the first-order forward difference, this feature is defined as:

$$|\Delta|^{(s)} = \frac{1}{N-1} \sum_{t=1}^{N-1} \left| x_{t+1}^{(s)} - x_t^{(s)} \right| \tag{3.4}$$

Thus, for each signal, four per-cycle features are produced: mean, standard deviation, skewness, and average magnitude of the derivative. In addition to the primary sensor signals, three cycle level signals were also appended as features: cycle_length, idle_time, and idle_count. Statistical features were therefore computed for a total of six included signals. With four statistical measures per signal, this resulted in

$(4 \times 6) + 3 = 27$ initially constructed per-cycle features.

As stated in the limitations, not all available signals were included in the statistical feature extraction. The signals DC current 2 and SOC 2 were excluded from the mean, standard deviation, skewness, and derivative magnitude calculations and were only used for comparative inter-engine analysis. Consequently, they do not contribute to the per-cycle feature vector used for clustering. In addition, an initial feature screening step was performed after feature computation. Features that did not provide informative variation was removed. A feature was excluded if it contained too few valid samples across cycles or if its values were constant or near constant, for example dominated by zeros. Such features do not contribute to cluster discrimination and may negatively affect distance based methods. Based on these criteria the three computed features in Table 3.2 were removed. After exclusion of these non informative features and the signals removed as per the limitations of the thesis, the remaining 24 features form the final per-cycle feature vectors used in the clustering analysis.

Table 3.2: Derived features removed due to low information content or insufficient variation.

Signal	**Derived feature**
Dc_diff	mean
SOC1	$\|\Delta\|$
SOC_diff	$\|\Delta\|$

3.3.1 Principal component analysis

Principal component analysis (PCA), was applied to reduce dimensionality and noise in feature space. The numeric features were assembled into a single vector and standardized, given a mean value, $\mu = 0$, and a standard deviation, $\sigma = 1$, so that larger scales did not generate any bias and dominate the analysis. The PCA was then fitted and the cumulative explained variance was added and inspected in order to distinguish the number of principal components that were relevant. A 95 percent variance threshold was chosen to retain the smallest number of principal components to ensure that most of the information was preserved while discarding unnecessary information.

3.3.2 Principal Feature Analysis

The representative features were extracted from the PCA by weighing each principal component(PC) dimension by the square root of its explained variance, $\sqrt{variance}$. This was done to ensure that components with greater influence, and therefore greater variance, have greater influence. Each feature is now defined as a single point in the transformed PC space. In order to categorize/group different points together K-Means was used. The number of clusters was chosen equal to the number of retained PCs, $k = PCs$. This choice is motivated by the interpretation of each PC as a dominant mode of variation in the data, and by prior work showing a close

relationship between clustering structure and the principal space of the data [21]. From each cluster the feature closest to the centroid was chosen as the representative feature.

3.4 Clustering

After the feature extraction and identification of the most relevant features, the next step is to group the data into clusters. These data clusters will be crucial for deeper understanding of the powertrain behavior and identifying what the different drive cycles represent. To achieve a satisfactory result, different clustering methods was evaluated on their performance in grouping the data in the most efficient and informative way.

3.4.1 Hierarchical Agglomerative Clustering

HAC was performed with ward's linkage to minimize the increase in cluster variance at each branching in the tree. A dendrogram was then constructed to visualize the branching of the drive cycles.

3.4.2 K-Means

K-Means clustering was applied to the final dataframe produced from the preprocessing and feature-scaling steps. The algorithm aims to partition the data into k clusters by minimizing the within-cluster sum of squared distances to the cluster centroids. To determine an appropriate number of clusters, k, an internal validation study was conducted using three commonly used metrics: the Silhouette Score, the Calinski–Harabasz Index, and the Davies–Bouldin Index. These metrics quantify cluster separation, compactness, and overall structure.

Based on the evaluation, k=2 was identified as the optimal number of clusters. K-Means was then executed with this value, and each cycle in the dataset was assigned to one of the two clusters. To further assess the stability and interpretability of the clustering solution, the resulting K-Means labels were compared against those produced by the K-Medoids algorithm using a cluster correspondence (confusion) matrix. This comparison was then used to evaluate how consistently both methods grouped the underlying data patterns.

3.4.3 K-Medoids

K-Medoids clustering was also applied to the same feature space to provide a robustness check against the centroid-sensitive nature of K-Means. Unlike K-Means, which uses the mean of points as the cluster center, K-Medoids selects actual data points (medoids) as cluster representatives, making it more resilient to noise and outliers. The same set of cluster validation metrics—Silhouette Score, Calinski–Harabasz Index, and Davies–Bouldin Index—were used to determine the optimal number of clusters.

Consistent with the findings from K-Means, k=2 was again selected as the most appropriate choice. The K-Medoids output was then compared to the K-Means results using a confusion matrix to assess alignment between the two algorithms and to verify the stability of the discovered structure in the data.

3.4.4 Convolutional Autoencoder

A CAE was built consisting of two main parts: an encoder and a decoder. The autoencoder based clustering is used as an exploratory complement to feature based clustering, with the aim of identifying finer grained temporal patterns. In figure 3.2 the structure of the CAE can be seen. The encoder contains two convolutional layers, each followed by a ReLU activation, and two max-pooling layers that progressively reduce the spatial resolution and extract higher level representations. The decoder mirrors this structure using two transposed convolutional layers to upsample the encoded representation, followed by a ReLU activation and a final Sigmoid activation to reconstruct the input signals.

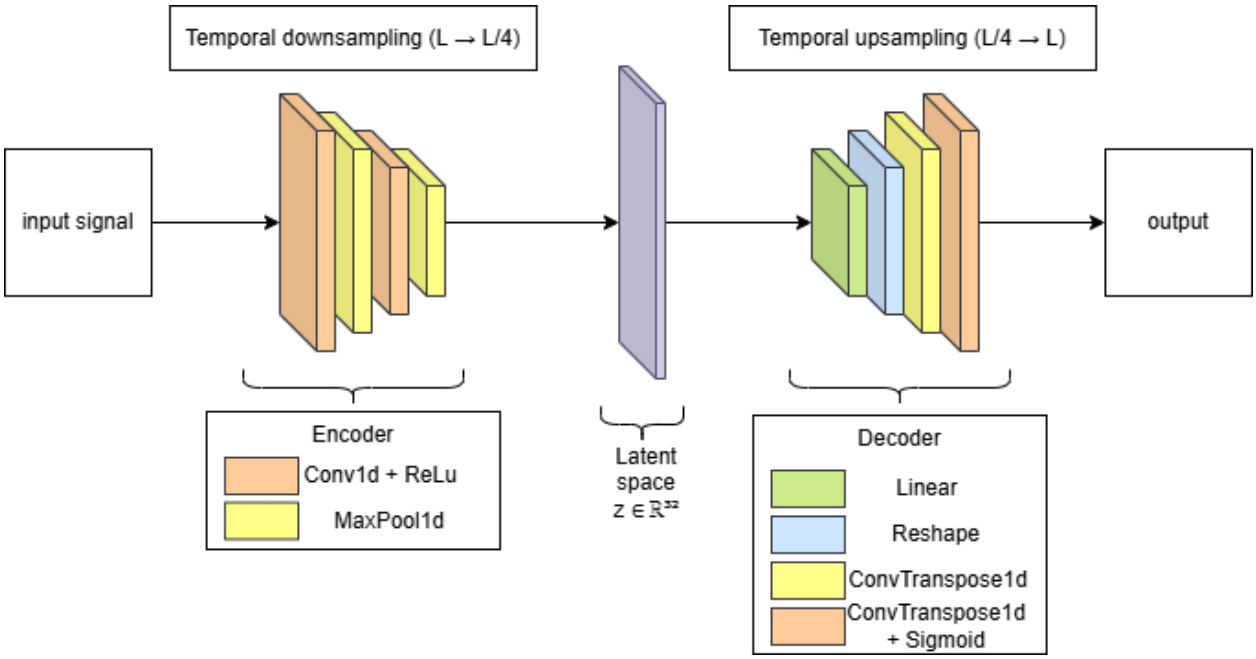


Figure 3.2: Structure of the convolutional auto encoder.

The CAE was trained on the cycle segmented time series data. During training, the learnable parameters of the network, specifically the weights and biases of the convolutional and transposed convolutional layers, were optimized using backpropagation to minimize the reconstruction loss between the input signals and their reconstructions. The resulting latent vectors represent compressed embeddings that capture the essential structure of each cycle. The dimensionality of the latent vector z was set to 32, chosen as a reasonable starting point based on the complexity of the input signals. This value was not systematically optimized and could be treated as a hyperparameter for further tuning. Training was performed for 150 epochs using a learning rate of 0.001. Training and validation losses were monitored throughout, and training was stopped once the loss curves converged to avoid overfitting. After training, K-Means clustering was applied to the latent vectors to discover groups of cycles. Standard clustering metrics were used to determine the optimal number of clusters, and K = 6 was selected. Finally the original signals were assessed and interpreted to evaluate the performance of the Auto encoders clustering.

3.5 Road cycle creation and simulations.

Once the drive cycles were defined, road cycle and vehicle simulations began. The road cycles served as the foundation for the vehicle simulations and results. The road cycles and vehicle simulations were created using Volvo Penta provided Matlab- and Simulink files. Figure 3.3 shows the overall structure of the simulations part of the project. Speed(km/h) signals, GPS position signals such as altitude, longitude and latitude, as well as timestamps are extracted from the representative and outlying drive cycles and then used to create a road cycle. This road cycle is fed into the simulations model where two different configurations of weight are simulated for each of the chosen cycles. As a result of the simulations, signals from both the engine and the battery can be analyzed.

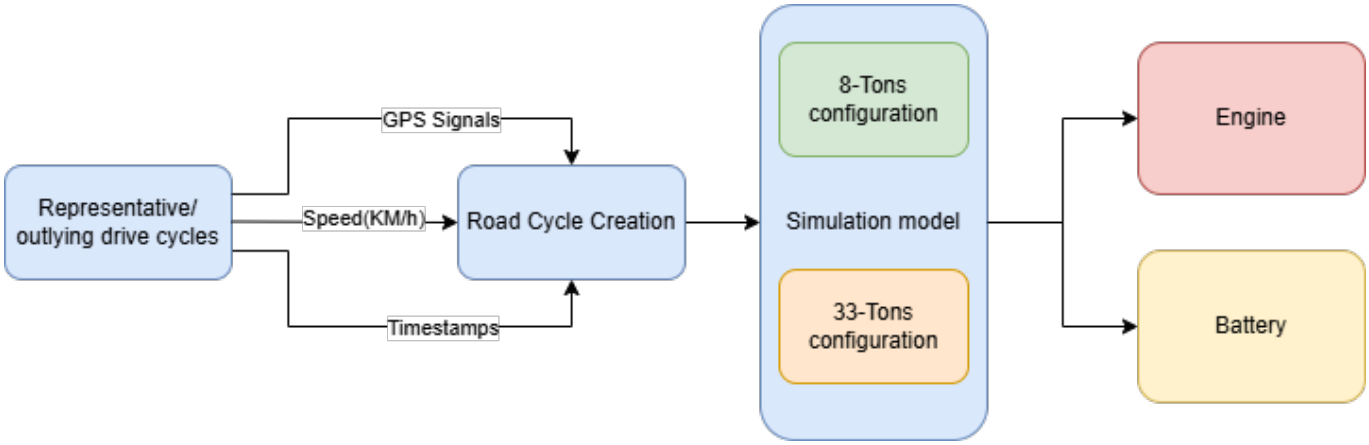


Figure 3.3: Flowchart of the methodology for the simulations.

Road cycle creation

The first step of the road cycle creation was to extract data from the clustered drive cycles. After the data extraction, the input data was divided into two categories. The first category being drive cycle-specific and containing timestamped measurements of: vehicle speed, longitudal, latitudinal and altitudinal position, acceleration and measurement duration. The second category of the selected input data contains general vehicle information such as: start-stop levels of SOC, gear ratios, wheel diameter, etc. The first category of input data was used for the road cycle creation and the second category of input data was used for the drive cycle simulations. Each category of the input data was organized as matrices enabling Matlab operations.

The road cycle was created to verify the speed profile, average speed, covered altitude and the driving route for the terminal tractor. To determine the road cycle speed the input speed data was filtered and resampled. The distances were computed by integration of the speed over the timestamps and then validated against the GPS-position data. The difference in distance indicates how accurate the simulated model is when compared to the raw input data. The difference arose due to the Matlab-scripts interpolation of the consecutive timestamps.

Speed profile and global treatment

During the creation of the road cycle a target set speed, V_{set}, was created. This speed is the output and represents a filtered version of the input speed. The applied treatment extended the acceleration and deceleration phases in the drive cycle. This made sure the powertrain had enough time to reach the correct speed levels during the drive cycle so that the road cycle could be used for simulation.

Following the generation of the output speed, V_{set}, a filter was applied to remove extended idle periods and interpolation values. The filter also smoothens transient spikes for the speed values. After this step the output speed profile contained the same properties as the original input speed data while being more stable, ensuring sufficient time for the powertrain to reach the desired speed levels for simulation.

Altitude, road slope and road cycle validation

After the output speed was generated, the altitude and road slope were investigated. The altitude was reconstructed by combining the atmospheric pressure measurements, GPS altitude and the external map elevation data. The final altitude profile contains smoother transitions and the road slope profile was derived from the output altitude with respect to the covered distance. The last step applied for the road cycle creation was to validate the speed. The output values were combined to create a new road cycle and altitude profile that was plotted against the original ones.

Simulations

Once the road cycles were created, confirming that as much as possible from the input data was conserved, terminal tractor simulation began using Volvo Penta developed models in both Matlab and Simulink. The set data from category two, such as start-stop levels of SOC, wheel diameter, etc. was now incorporated into the model. The majority of the input values were set to predetermined values after consulting with Volvo Penta domain experts. These values were selected to represent the most basic/average/standard vehicle operating conditions.

The exception was the SOC limits and the vehicle mass. The SOC limits were changed to be at 80% as fully charged when initiating the drive cycles and 20% was considered fully depleted. The reasoning behind this parameter selection was as follows. Firstly, the filtered data didn't incorporate charging events, only regenerative charging, meaning that high SOC levels wouldn't appear. The second reason was that after analyzing the end-values of the SOC, below 20%-values rarely occurs. The third reason was that after consulting Volvo Penta domain experts, the 80%-20% SOC operating window is recommended for the examined terminal tractor and therefore should be an expected operating window. The last reason supporting the selected SOC limits was that the input data doesn't include vehicle mass, resulting in unpredictable energy consumption during the drive cycle simulations.Therefore by allowing operation within the 80%-20% SOC range the most accurate representation, while simultaneously removing the battery as a limiting factor, was utilized and therefore all simulations were performed within the 80%-20% SOC operating window.

The second parameter chosen to vary was the vehicle mass. Here the choice was made to use a configuration of 8-ton and 33-ton. The reason behind the selected vehicle masses was motivated by the maximum and minimum weight. When the terminal tractor operates without any attachments, such as containers, the sole vehicle weight is 8-ton and the maximum cargo it is designed to move, in the standard vehicle configuration, is of 25-ton. So by selecting 8- and 33-ton as the vehicle masses for the simulations, the lightest, as well as the heaviest, possible vehicle configuration was examined.

Once the road cycles were created and vehicle parameters determined, the cluster generated drive cycles were simulated. Figure 3.4, highlights the combination of the set value parameters, SOC levels and vehicle mass used for one simulation. Once a drive cycle simulation was done the process was repeated until all drive cycles were simulated.

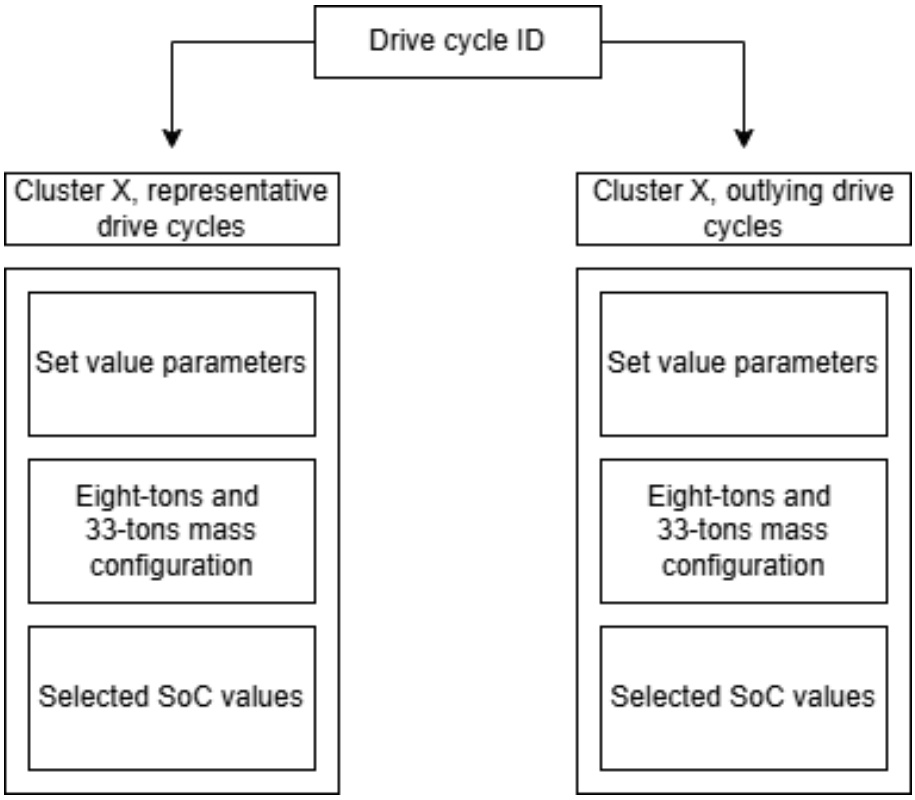


Figure 3.4: Description of the simulation process for the identified drive cycles.

4
Results

This chapter presents the results found during the work of this thesis. The chapter is divided into the following section: *Cycle definition, Feature selection, Clustering and Simulations* with a discussion at the end of each part.

4.1 Cycle definition

The segmentation procedure as described in Section 3.2 produced a variety of cycles with differing lengths, reflecting the natural variability of operational behavior. As illustrated in Figure 4.1, the distribution is strongly right skewed, with a large concentration of shorter cycles and a long tail of extended operations. Cycles shorter than 15 minutes were removed because such fragments were deemed insufficient to describe meaningful engine behavior.

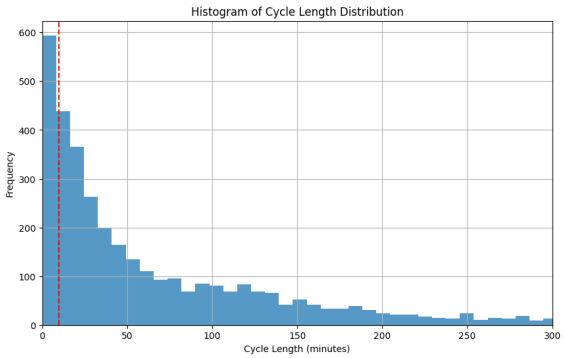


Figure 4.1: Plot showing the length of each cycle in minutes and the frequency at which it appears. With the threshold of 15 minutes shown by the dotted vertical line.

After applying this threshold, 3295 cycles remained, while 1397 were discarded. The preserved set cover most of the distribution of cycles. This means the dataset still includes short, medium and long duty cycles, and is fit for further analysis with clustering. The filtered dataset therefore provides a more stable foundation for extracting behavioural structure in later stages of the analysis, reducing noise from smaller cycles that would otherwise distort both feature and distance metrics.

Discussion

While the 15 minute threshold effectively removes cycles that lack sufficient behavioural significance, it also introduces a natural trade off. Some genuine but brief operational events may be excluded, which could bias the dataset toward longer and more complex duty patterns. For the purposes of this study, the benefits outweigh the loss, as the clustering models rely on consistent and interpretable feature distributions. Nevertheless, future work could explore adaptive thresholds, probabilistic segmentation confidence, or weighting schemes that preserve short cycles without letting them dominate the feature landscape.

4.2 Feature selection

The feature selection combined PCA driven dimensionality reduction with a subsequent PFA step to identify a representative set of variables for clustering. After preprocessing, PCA was applied to the full feature matrix to examine the global variance structure. Figure 4.2 displays the cumulative explained variance analysis which showed that 17 out of the original 24 principal components captured more than 95% of the total variance.

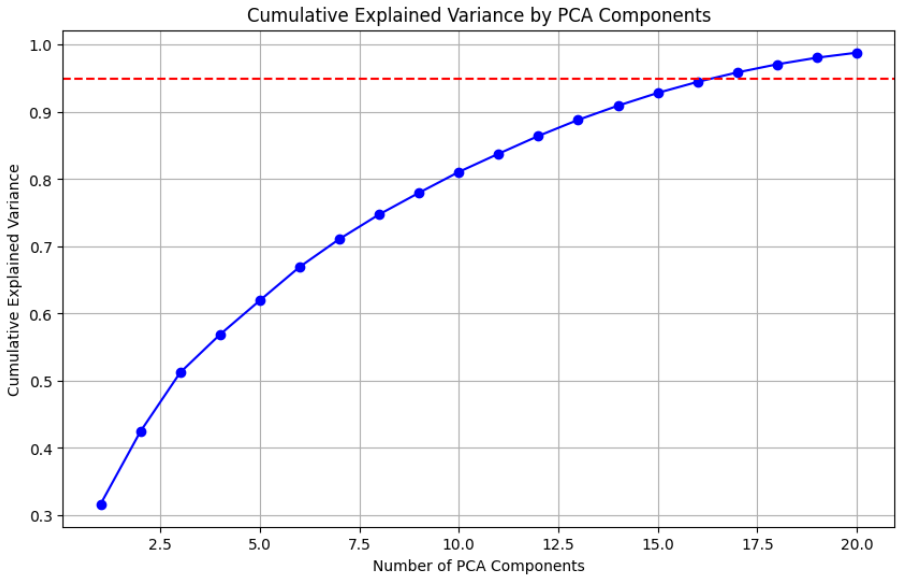


Figure 4.2: The cumulative explained variance with a threshold at 95% explained variance.

This structure is further illustrated in Figure 4.3, which shows the projection of all samples onto the first three principal components. The distribution reveals a dense core with elongated directions of variance, confirming that the data lie on a low dimensional manifold rather than occupying the full high dimensional feature space uniformly.

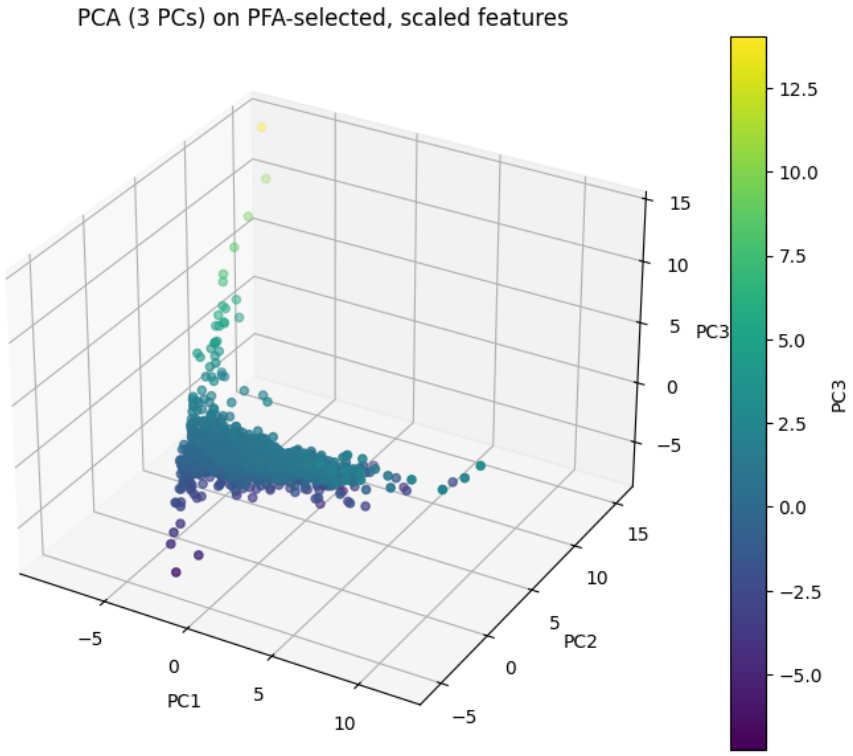


Figure 4.3: Projection of samples onto the first three principal components computed from PFA-selected, scaled features. The color indicates the third principal component value.

Based on the selected number of components, feature loadings were transformed into PCA space, weighted by component importance, and grouped using K-Means clustering. Each cluster thus formed a set of features with similar behavior across the dominant principal components. For each cluster, the feature closest to the centroid in loading space was selected as the representative variable.

Table 4.1 displays all features used in PCA/PFA, their centroid distances and cluster id, With the features selected by PFA from each cluster highlighted in bold. As can be seen the distances for each feature are close to equal, even though some discrepancies can be seen. A large subset of features, including Torque_skewness, Speed_mean, SOC1_mean, SOC1_σ, DcCurrent_diff_skewness, and DcCurrent_diff_σ, exhibit zero or near-zero distance to their respective cluster centroids. Even though the differences are small, a tiny difference in distance to a specific cluster centroid is enough to be excluded by the PFA.

Table 4.1: Full feature list sorted by cluster ID. Features selected by PFA are shown in bold, with approximate distances for visibility.

Feature name	Distance to centroid	Cluster ID
Torque_σ	1.57×10^{-16}	0
Speed$\lvert\Delta\rvert$	1.11×10^{-16}	1
SOC1_mean	2.22×10^{-16}	2
DcCurrent_diff_σ	0	3
Torque_skewness	2.0012	4
Dc1_skewness	2.0012	4
Torque_mean	0.9026	5
Dc1_mean	1.5816	5
DcCurrent_diff$\lvert\Delta\rvert$	1.4411	5
SOC1_σ	1.92×10^{-16}	6
SOC_diff_mean	2.0440	7
SOC_diff_skewness	2.0440	7
SOC_diff_σ	1.11×10^{-16}	8
idle_time_minutes_mean	1.8529	9
idle_count_mean	1.8529	9
Dc1$\lvert\Delta\rvert$	0	10
Speed_mean	0	11
SOC1_skewness	2.22×10^{-16}	12
cycle_length_seconds_mean	2.22×10^{-16}	13
Torque$\lvert\Delta\rvert$	1.7480	14
Dc1_σ	1.7599	14
Speed_σ	2.2833	14
Speed_skewness	0	15
DcCurrent_diff_skewness	1.11×10^{-16}	16

A second group of representative features, such as Speed_skewness and cycle_length_seconds_mean, also shows marginally nonzero centroid distances, as reported in table 4.1. The extremely small magnitude of these deviations suggests that these features remain tightly aligned with their cluster structure, and their slightly lower rank reflects numerical sensitivity rather than a loss of representability.

In contrast, features such as *Torque_mean*, *Torque_σ*, and *idle_time_minutes_mean* exhibit comparatively larger distances to their cluster centroids. As shown in table 4.1, these features originate from clusters with weaker internal structure, indicating less tightly correlated feature sets. Although they are still selected as representatives, their reduced centrality suggests that the physical processes they describe are more variable.

Overall, the PFA procedure results in a stable and interpretable set of representative features that summarize the dominant variable patterns in the dataset while reducing redundancy. The combination of highly central features from well structured clusters and a smaller number of representative features further from more diffuse clusters ensures coverage of broader operational behavior.

4.2.1 Discussion Feature selection

The feature selection results highlights several important properties of the dataset and the methodological choices. Firstly, clustering observed in PCA loading space confirms that many raw features describe highly overlapping physical behavior. The dominance of skewness, mean, and derivative based features, among the top ranked features suggests that behavioural asymmetries, load transitions, and state of charge dynamics are among the most informative descriptors of cycle behavior.

The appearance of tightly formed PFA clusters indicates that the selected features capture well structured subspaces of operational variability. This strengthens the interpretability of later clustering analysis, as each selected feature represents a broader class of correlated variables rather than an isolated measure.

The weaker cluster features: *Torque_σ*, *Torque_mean*, and *idle_time_minutes_mean* and their low PFA scores imply that the physical processes they describe may be more variable, less cleanly correlate to the other metrics, or is influenced by noise or operational artifacts. Their inclusion as representatives ensures that these broader but less coherent behaviours remain covered.

Finally, the combination of PCA and PFA provides interpretable features that corresponds heavily to the internal structure of the analysed signals. PCA alone compresses information into abstract components, while PFA restores physical meaning by mapping these components back to measurable features. The resulting feature set is therefore both compact and grounded in relevant signals, which is essential for clustering, profiling, and cycle interpretation.

4.3 Clustering

This section evaluates the results obtained from clustering with the aim of identifying distinct operating regimes within the engine cycle data. Each clustering method is assessed using quantitative validity metrics, low dimensional visualizations, and an examination of representative and outlying cycles to ensure both statistical and physical interpretability. For the sake of visibility in the tables, the anonymisation of terminal tractor names has in some instances been shortened i.e Brave_Hawk_6 = BH_6... and so on.

4.3.1 Hierarchical Agglomerative Clustering

Figure 4.4 summarizes three standard cluster validity metrics across different values of k. The silhouette score peaks at $k = 2$, while both the Calinski–Harabasz and Davies–Bouldin indices similarly indicate that increasing the number of clusters does not lead to clearer separation or improved density. Together, these metrics support the choice of $k = 2$.

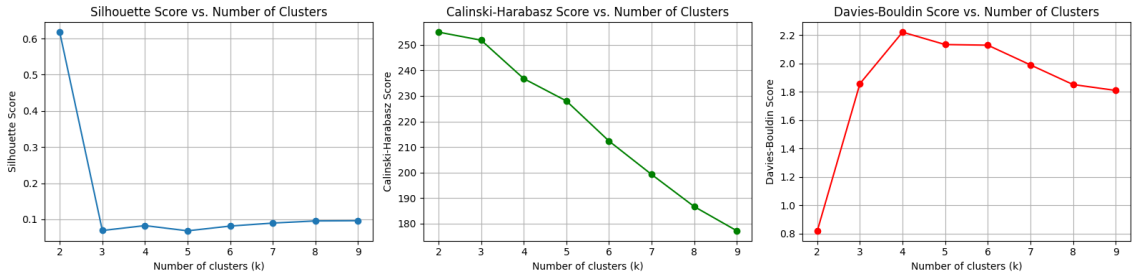


Figure 4.4: Plots describing the three standard metrics, silhouette score, Calinski-Harabasz and Davies-Bouldin indices to assess clustering quality for HAC.

Figure 4.5 shows the dendrogram produced by HAC. The structure reveals two dominant branches separated at a distance threshold of 80, which provides a natural division of the dataset into two primary clusters in accordance with the clustering metrics above. The relatively balanced merge heights within each branch indicate similar internal groups with comparable variabilty.

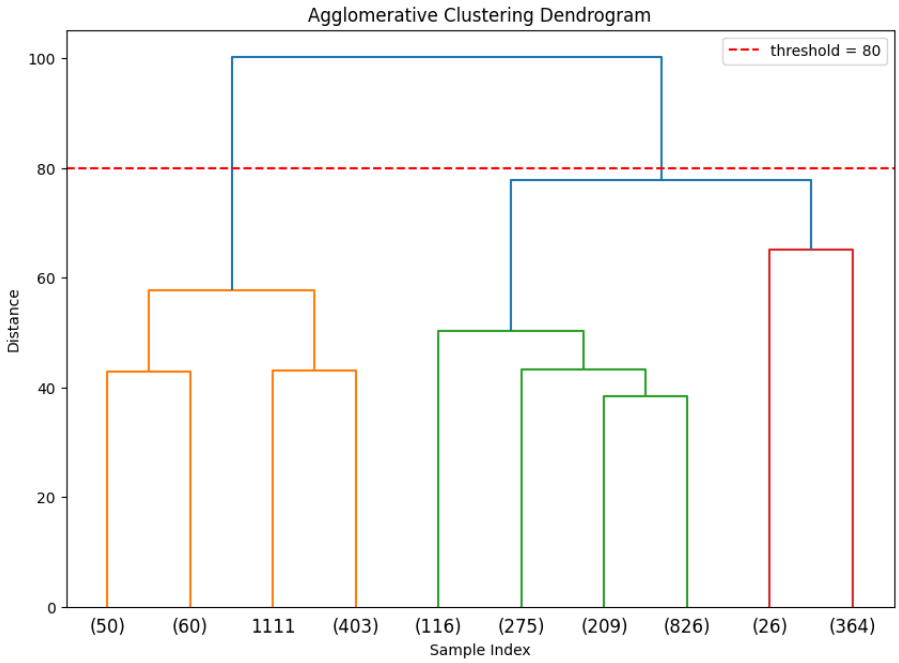


Figure 4.5: Z-Dendrogram of the HAC, with the distance threshold at 80.

Figure 4.6 visualizes the resulting clusters in the PCA-projected feature space. The two clusters form elongated regions along the principal axes, with clearly separated

centroids. The elongated cluster shapes indicate that separation is driven primarily by a subset of dominant features rather than uniform variation across all dimensions. The transition around zero along the first principal component suggests that the primary mode of variation corresponds to changes in dynamic behavior rather than absolute operating level. A few samples appear as outliers, confirming the presence of cycles that deviate from the main operational patterns.

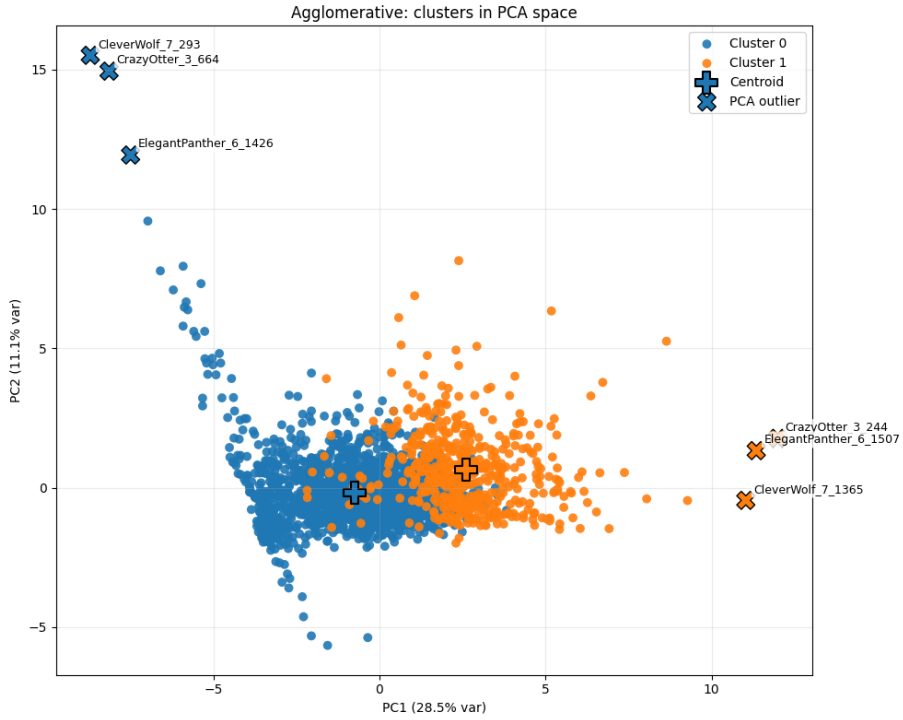


Figure 4.6: Clusters visualized in PCA space, with Cluster 0 containing 1816 cycles and Cluster 1 containing 514 cycles for HAC.

Figure 4.7 presents pairwise feature relationships for the four most important features. Skewness related attributes which capture asymmetric behavior in torque, speed and current signals. The density distributions highlight that Cluster 0 tends to exhibit more skewed feature values, whereas Cluster 1 has more variance in the standard deviation features. This separation indicates that the features contribute to the separation of the clusters into two operational modes, Cluster 1 with more varied drive cycles and Cluster 0 with a more stable operating mode.

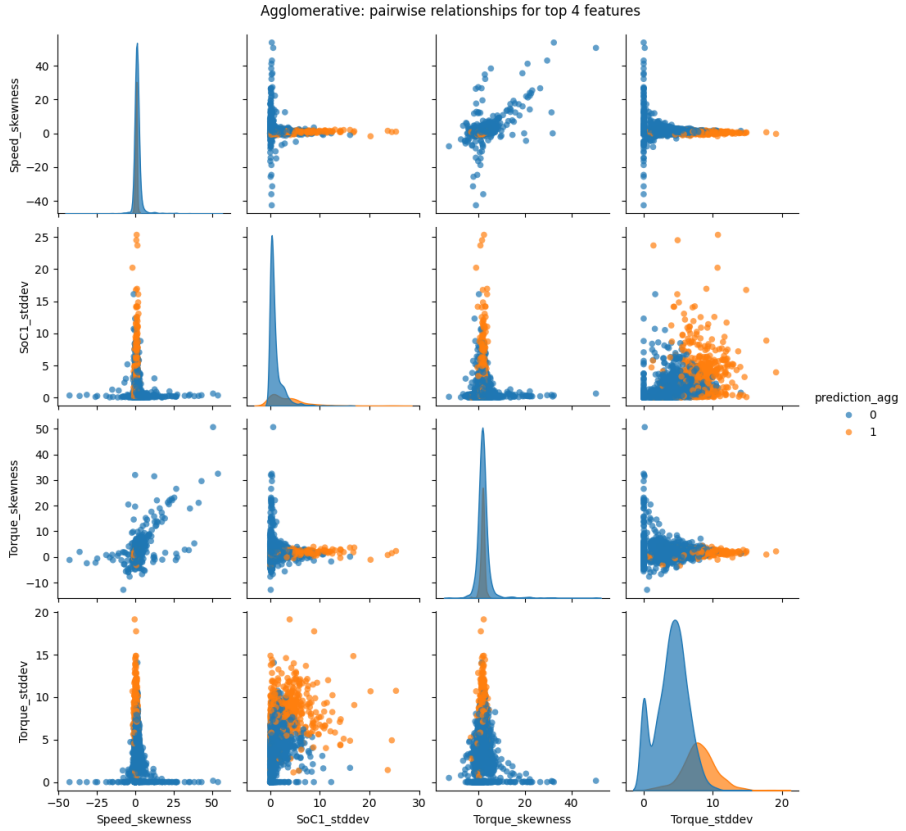


Figure 4.7: Pairplot of the 4 most representative features in feature space.

Table 4.2 showcases the three most representative cycles for each cluster that was selected by the algorithm.

Table 4.2: Selected representative cycles using HAC.

Cluster	Cycle ID
0	ElegantPanther_6_1205
0	ElegantPanther_6_390
0	ElegantPanther_6_562
1	CleverWolf_7_1291
1	CleverWolf_7_1740
1	CleverWolf_7_872

Table 4.3 shows that the representative cycles in Cluster 0 exhibit relatively consistent and low feature values across all selected attributes. Speed and torque skewness remain within a narrow range, while derivative based features such as Speed $|\Delta|$ and Torque $|\Delta|$ indicate moderate but controlled dynamic behavior. This consistency supports the interpretation of Cluster 0 as a group of varied but stable driving cycles, where asymmetry in signals is present without extreme fluctuations.

Table 4.3: Feature values extracted from representative cycles in Cluster 0 using HAC.

Feature	EP__6__1205	EP__6__390	EP__6__562
Speed skew	1.717	1.726	1.525
SoC1 σ	1.289	1.416	1.033
Torque skew	2.961	2.497	1.448
DcCurrent diff skew	-0.498	0.809	3.957
Torque mean	0.911	1.004	1.040
Torque $\lvert\Delta\rvert$	1.137	1.069	0.816
DcCurrent diff σ	4.417	5.141	4.850
Speed $\lvert\Delta\rvert$	90.270	81.424	81.804

In contrast, Table 4.4 highlights that representative cycles in Cluster 1 display substantially higher values for standard deviation and derivative based features. Increased *SoC1_*σ, *DcCurrent_diff_*σ, and *Speed_*$\lvert\Delta\rvert$ indicate stronger dynamic excitation and larger operating variability. These patterns suggest that Cluster 1 corresponds to a more aggressive or transient operating regime, characterized by frequent changes in load and speed.

Table 4.4: Feature values extracted from representative cycles in Cluster 1 using HAC.

Feature	CW__7__1291	CW__7__1740	CW__7__872
Speed skew	0.760	0.632	0.641
SoC1 σ	4.711	4.872	3.822
Torque skew	1.876	1.683	1.523
DcCurrent diff skew	1.074	-0.757	0.100
Torque mean	2.300	2.179	2.562
Torque $\lvert\Delta\rvert$	2.816	2.268	3.217
DcCurrent diff σ	9.354	7.289	7.447
Speed $\lvert\Delta\rvert$	166.528	174.031	167.998

Table 4.5 displays the three most outlying cycles for each cluster selected by the algorithm.

Table 4.5: Selected outlier cycles using HAC.

Cluster	Cycle ID
0	BraveHawk_4_196
0	CrazyOtter_3_54
0	ElegantPanther_6_906
1	CleverWolf_7_470
1	CleverWolf_7_559
1	CleverWolf_7_129

The outlying cycles in Cluster 0 shown in Table 4.6 exhibit feature values that deviate strongly from the representative patterns. In particular, EP_6_906 displays extreme torque skewness and highly negative *DcCurrent_diff_skewness*, indicating pronounced asymmetry and irregular current behavior. These deviations suggest rare operational scenarios such as abrupt load transitions or atypical drive events that are not representative of the cluster as a whole.

Table 4.6: Feature values extracted from outlying cycles in Cluster 0.

Feature	**BH_4_196**	**CO_3_54**	**EP_6_906**
Speed skew	1.633	0.826	4.854
SoC1 σ	3.204	0.618	0.800
Torque skew	1.618	1.657	8.723
DcCurrent diff skew	2.936	3.509	-21.850
Torque mean	1.315	1.167	0.231
Torque $\lvert\Delta\rvert$	0.753	1.045	0.191
DcCurrent diff σ	1.263	4.576	2.564
Speed $\lvert\Delta\rvert$	76.986	109.932	15.965

Table 4.8 demonstrates that outliers in Cluster 1 are primarily characterized by extreme variability rather than skewness alone. High values of *Speed_*$\lvert\Delta\rvert$ and *DcCurrent_diff_*σ indicate strong transient behavior, while elevated *Torque_*$\lvert\Delta\rvert$ suggests aggressive torque modulation. These cycles likely correspond to highly demanding operating conditions that exceed the typical behavior even within the more dynamic Cluster 1.

Table 4.7: Feature values extracted from outlying cycles in Cluster 1 using HAC.

Feature	**CW_7_470**	**CW_7_559**	**CW_7_129**
Speed skew	0.877	0.184	0.723
SoC1 σ	8.747	0.429	0.261
Torque skew	0.383	1.143	1.302
DcCurrent diff skew	-2.112	0.453	6.093
Torque mean	0.777	5.503	3.076
Torque $\lvert\Delta\rvert$	2.473	4.203	3.714
DcCurrent diff σ	7.042	14.079	8.928
Speed $\lvert\Delta\rvert$	190.599	300.273	233.281

Discussion Hierarchical Agglomerative Clustering

HAC proved well suited for this type of data due to its ability to capture non spherical and hierarchically structured patterns without imposing assumptions on cluster geometry. Hierarchical clustering is widely used in exploratory data analysis due to its ability to reveal nested cluster structure through dendrogram representations [39]. The dendrogram provides an interpretable representation of how individual drive cycles group together, revealing a clear separation into two dominant operational

regimes. However, it is important to note that alternative signal representations or feature selections may separate the data more clearly or reveal additional subgroups.

The resulting clusters are not only supported by quantitative validity metrics but also by consistent physical interpretations derived from feature behavior. Representative cycles within each cluster display coherent patterns, while outliers highlight deviations. The consistency between quantitative validity metrics and physically interpretable feature patterns observed in this study supports the suitability of HAC for exploratory analysis of complex engine cycle behavior.

4.3.2 K-Means Clustering

Figure 4.8 presents the cluster validity metrics across different values of k. The silhouette score decreases steadily after $k = 2$, and both the Calinski–Harabasz and Davies–Bouldin indices showcase deterioration with higher cluster counts. These trends suggest that increasing the number of clusters does not yield better separations, reinforcing that the data naturally supports a $k = 2$ structure.

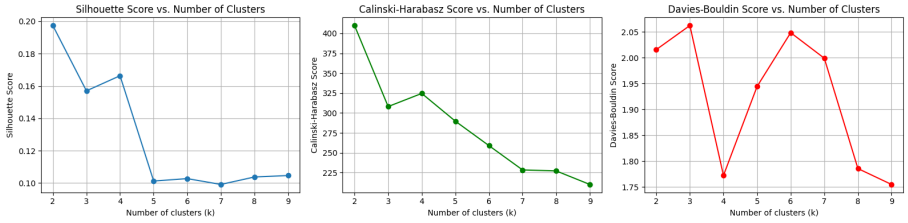


Figure 4.8: Plots describing the three standard metrics, silhouette score, Calinski-Harabasz and Davies-Bouldin indices to assess clustering quality for K-Means.

Figure 4.9 shows pairwise relationships for the four most influential features. The two clusters separate most clearly along *Torque_σ* and the skewness based attributes. Cluster 1 consistently exhibits higher variance in standard deviation features, while Cluster 0 remains closer to the central data with wider spread in skewness related features. This indicates that the clusters capture distinct operational patterns, one representing stable operation and the other characterized by stronger fluctuations.

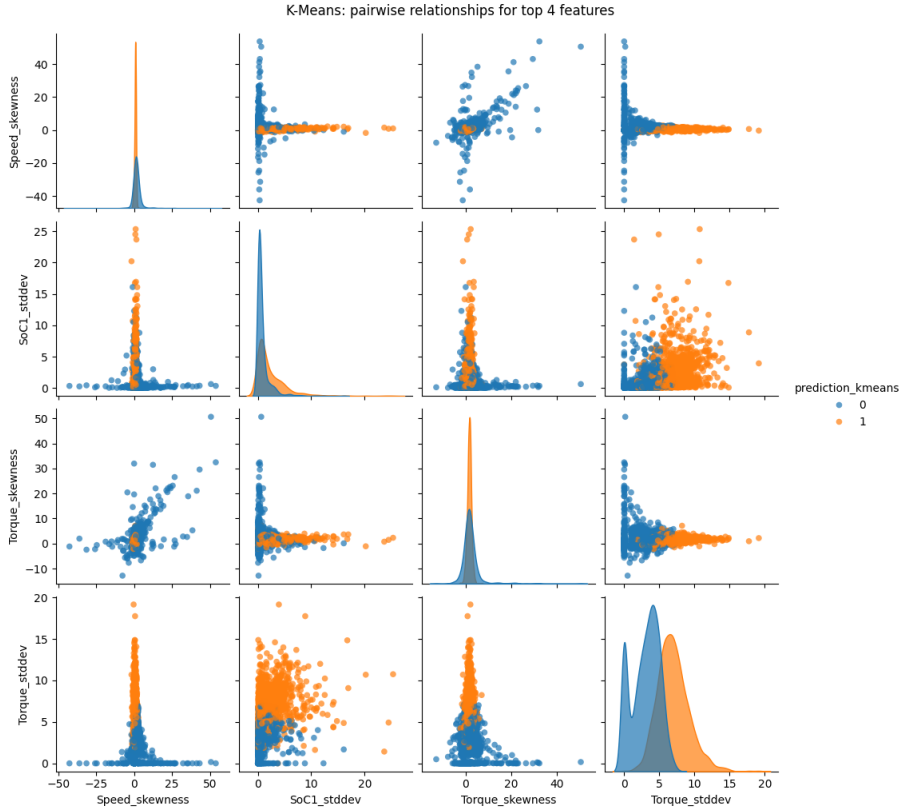


Figure 4.9: Pairplot of the four most representative features in feature space for K-Means clustering.

Figure 4.10 visualizes the clusters in PCA space. The two groups form elongated regions with well separated centroids, similarly to the skewness driven structure seen in the feature plots. A small set of points lies far from the main distribution, confirming the presence of a few outlier cycles that the model isolates.

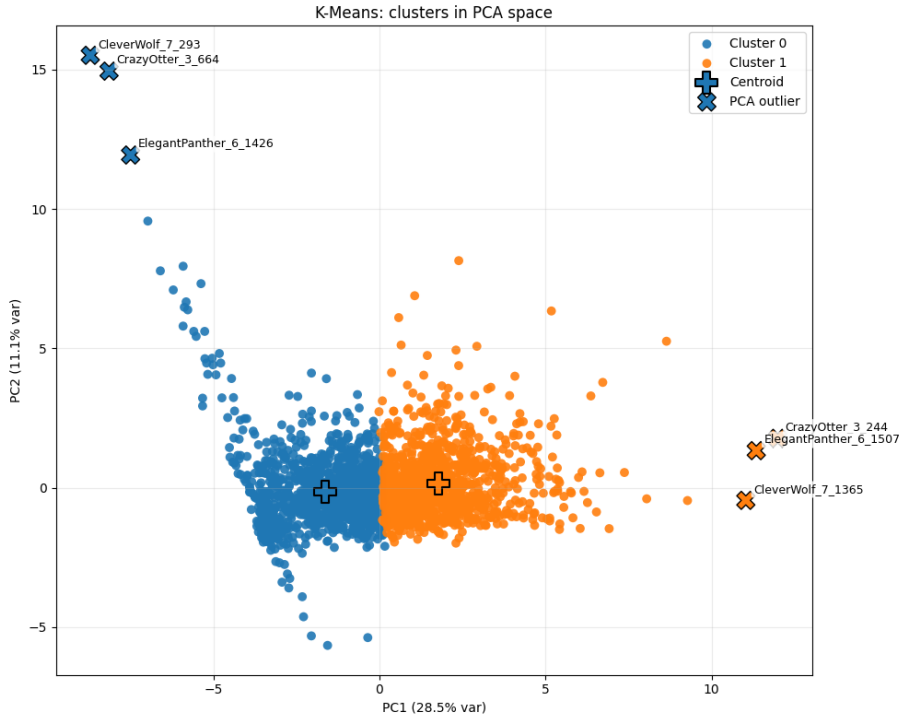


Figure 4.10: Clusters visualized in PCA space, with Cluster 0 containing 1219 cycles and Cluster 1 containing 1111 cycles for K-Means clustering.

Figure 4.11 highlights the distribution of *Torque_σ* across clusters. Cluster 1 shows increased variability and a wider tail of extreme values, confirming that variability in torque is a dominant factor driving the separation of the two groups. Cluster 0 remains tightly bounded, supporting the interpretation that it represents more consistent and stable driving cycles.

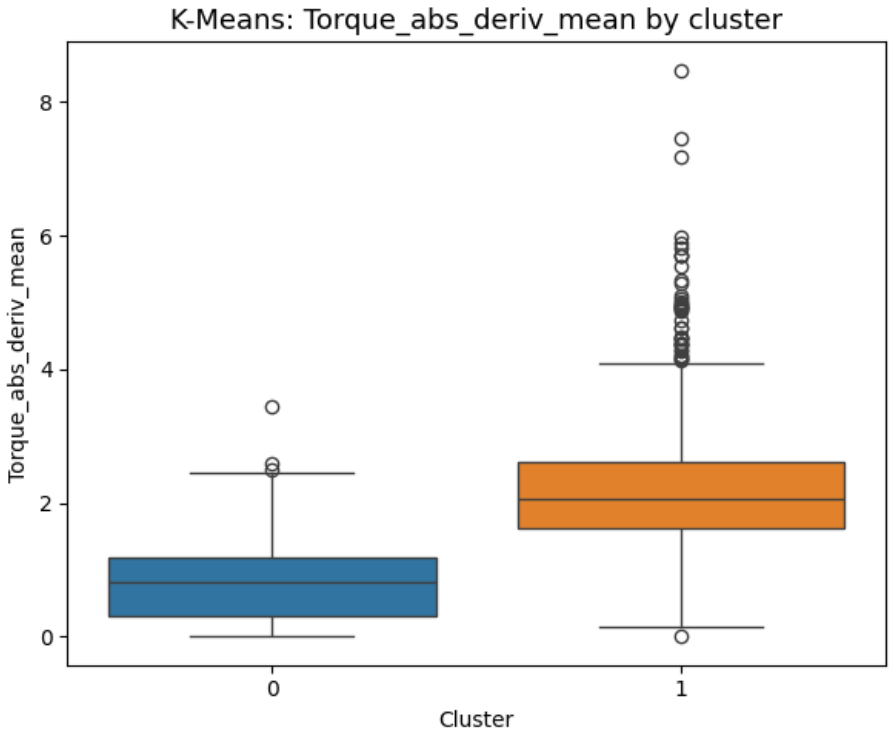


Figure 4.11: Boxplot of the torque absolute derivative mean value distribution by cluster for K-Means.

Table 4.8 showcases the selected representative drive cycles extracted from the K-Means clustering algorithm.

Table 4.8: Selected representative cycles using K-Means clustering.

Cluster	Cycle ID
0	CrazyOtter_3_487
0	BraveHawk_4_200
0	CleverWolf_7_120
1	CleverWolf_7_661
1	CleverWolf_7_135
1	CleverWolf_7_1548

The representative cycles for Cluster 0 selected by K-Means exhibit tightly grouped feature values across all examined attributes. Torque and speed skewness remain moderate, while standard deviation and derivative based features indicate controlled dynamic behavior. This suggests that Cluster 0 primarily captures stable driving cycles with limited transient excitation. Although asymmetry is present in some signals, the absence of extreme values indicates that these cycles reflect typical operational conditions rather than edge cases.

Table 4.9: Feature values extracted from representative cycles in Cluster 0 using K-Means clustering.

Feature	CO_3_487	BH_4_200	CW_7_120
Speed skew	1.868	1.829	1.693
SoC1 σ	0.359	1.189	0.179
Torque skew	1.802	2.550	2.298
DcCurrent diff skew	3.260	4.328	3.126
Torque mean	0.246	0.559	0.697
Torque $\lvert\Delta\rvert$	0.943	0.684	0.860
DcCurrent diff σ	2.227	2.112	2.472
Speed $\lvert\Delta\rvert$	85.315	60.090	74.582

In contrast, the representative cycles in Cluster 1 show consistently higher values for variability related features, particularly *Torque_*σ, *DcCurrent_diff_*σ, and *Speed_*$\lvert\Delta\rvert$. These elevated values indicate increased fluctuation in load, speed, and current, pointing toward a more demanding operating regime. Compared to Cluster 0, these cycles demonstrate stronger dynamic excitation and more frequent changes in operating conditions, supporting the interpretation of Cluster 1 as a high variability cluster.

Table 4.10: Feature values extracted from representative cycles in Cluster 1 using K-Means clustering.

Feature	CW_7_661	CW_7_135	CW_7_1548
Speed skew	0.611	0.862	1.142
SoC1 σ	2.424	2.184	2.616
Torque skew	1.432	1.499	1.506
DcCurrent diff skew	-0.223	3.146	1.760
Torque mean	1.729	1.970	1.722
Torque $\lvert\Delta\rvert$	2.560	2.186	2.099
DcCurrent diff σ	6.621	7.394	5.387
Speed $\lvert\Delta\rvert$	189.931	152.019	142.876

Table 4.11 displays the outlying drive cycles extracted from the K-Means algorithm.

Table 4.11: Selected outlier cycles using K-Means clustering.

Cluster	Cycle ID
0	CleverWolf_7_1466
0	ElegantPanther_6_496
0	ElegantPanther_6_317
1	ElegantPanther_6_1350
1	CrazyOtter_3_473
1	CleverWolf_7_861

The outlying cycles in Cluster 0 identified by K-Means exhibit feature values that deviate substantially from the representative patterns. In particular, elevated skewness and extreme DcCurrent_ diff_skewness indicate irregular and asymmetric behavior in the electrical signals. These cycles likely correspond to atypical driving scenarios such as abrupt load changes or irregular transient events, which are not representative of the dominant stable operating mode captured by the cluster

Table 4.12: Feature values extracted from outlying cycles in Cluster 0 using K-Means clustering.

Feature	**CW_7_1466**	**EP_6_496**	**EP_6_317**
Speed skew	2.446	1.772	7.742
SoC1 σ	5.820	5.985	0.490
Torque skew	2.789	1.787	14.523
DcCurrent diff skew	2.142	4.931	-3.192
Torque mean	0.417	1.197	0.102
Torque $\lvert\Delta\rvert$	0.856	0.854	0.093
DcCurrent diff σ	4.799	4.531	10.614
Speed $\lvert\Delta\rvert$	62.078	72.602	14.208

Outliers in Cluster 1 are characterized by exceptionally high variability and derivative values. Extremely large Speed $\lvert\Delta\rvert$ and Torque $\lvert\Delta\rvert$ suggest aggressive driving patterns with rapid changes in operating conditions. These cycles represent extreme operational demands that exceed even the typical behavior observed in the dynamic Cluster 1, highlighting rare but physically meaningful driving scenarios.

Table 4.13: Feature values extracted from outlying cycles in Cluster 1 using K-Means clustering.

Feature	**EP_6_1350**	**CO_3_473**	**CW_7_861**
Speed skew	-0.864	0.311	0.848
SoC1 σ	0.280	0.378	12.528
Torque skew	1.612	1.748	1.320
DcCurrent diff skew	-6.075	-2.398	0.444
Torque mean	3.200	3.248	2.170
Torque $\lvert\Delta\rvert$	1.987	5.704	2.120
DcCurrent diff σ	9.733	6.452	6.801
Speed $\lvert\Delta\rvert$	190.865	299.118	153.212

Discussion K-Means

Overall, K-Means and HAC produce consistent behavioral interpretations of the two clusters. In both methods, Cluster 0 is characterized by lower variability and more stable signal behavior, while the other is associated with higher variance and stronger transient dynamics driven by skewness and derivative related features. The outlier cycles identified by K-Means also exhibit the same type of high variability

and extreme derivative patterns as those observed with HAC. The main differences occur for cycles near cluster boundaries, where assignments may vary due to the different clustering principles. This agreement in cluster characteristics indicates that the identified operational modes are not method-specific but reflect underlying structure in the feature space.

4.3.3 K-Medoids Clustering

Table 4.17 showcases the chosen representative cycles for the two clusters using K-Medoids clustering. Further K-Medoids results can be seen in Appendix A due to its similar nature to K-Means.

Table 4.14: Selected representative cycles using K-Medoids clustering.

Cluster	Cycle ID
0	ElegantPanther_6_604
0	ElegantPanther_6_886
0	ElegantPanther_6_429
1	CleverWolf_7_661
1	CleverWolf_7_135
1	CleverWolf_7_271

The representative cycles selected by K-Medoids for Cluster 0 show strong internal consistency across all examined features. Compared to K-Means, the medoid based selection emphasizes cycles that are centrally located within the cluster rather than relying on averaged behavior. The resulting feature values indicate moderate skewness and controlled variability, reinforcing the interpretation of Cluster 0 as a stable operational regime with limited transient excitation.

Table 4.15: Feature values extracted from representative cycles in Cluster 0 using K-Medoids clustering.

Feature	EP_6_604	EP_6_886	EP_6_429
Speed skew	2.488	1.893	1.949
SoC1 σ	0.759	1.107	1.138
Torque skew	1.535	2.774	2.314
DcCurrent diff skew	-0.966	-1.491	5.958
Torque mean	0.419	0.662	0.659
Torque $\lvert\Delta\rvert$	0.805	0.761	0.872
DcCurrent diff σ	4.212	4.992	4.205
Speed $\lvert\Delta\rvert$	66.272	59.077	75.832

Representative cycles in Cluster 1 selected by K-Medoids display elevated variability and derivative based features, similar to the patterns observed with K-Means and HAC. High values of Speed $|\Delta|$, Torque $|\Delta|$, and DcCurrent_diff_σ indicate rapid changes in operating conditions and increased electrical and mechanical stress. The

consistency of these patterns across clustering methods strengthens confidence in the existence of a distinct high variability operating regime.

Table 4.16: Feature values extracted from representative cycles in Cluster 1 using K-Medoids clustering.

Feature	CW_7_661	CW_7_135	CW_7_271
Speed skew	0.611	0.862	0.964
SoC1 σ	2.424	2.184	2.929
Torque skew	1.432	1.499	2.097
DcCurrent diff skew	-0.223	3.146	-0.610
Torque mean	1.729	1.970	2.115
Torque $\lvert\Delta\rvert$	2.560	2.186	2.458
DcCurrent diff σ	6.621	7.394	7.695
Speed $\lvert\Delta\rvert$	189.931	152.019	162.288

Table 4.17 showcases the outlier cycles selected by the algorithm.

Table 4.17: Selected outlier cycles using K-Medoids clustering.

Cluster	Cycle ID
0	BraveHawk_4_196
0	ElegantPanther_6_906
0	CleverWolf_7_1722
1	ElegantPanther_6_331
1	CleverWolf_7_1207
1	BraveHawk_4_217

Table 4.15 showcases that the outlying cycles identified in Cluster 0 by K-Medoids exhibit pronounced deviations in skewness and derivative features, further confirming the results from previous algorithms.

Table 4.18: Feature values extracted from outlying cycles in Cluster 0 using K-Medoids clustering.

Feature	BH_4_196	EP_6_906	CW_7_1722
Speed skew	1.633	4.854	1.273
SoC1 σ	3.204	0.800	1.045
Torque skew	1.618	8.723	2.182
DcCurrent diff skew	2.936	-21.850	1.829
Torque mean	1.315	0.231	1.489
Torque $\lvert\Delta\rvert$	0.753	0.191	1.749
DcCurrent diff σ	1.263	2.564	5.975
Speed $\lvert\Delta\rvert$	76.986	15.965	92.594

Table 4.19 display that outliers in Cluster 1 selected by K-Medoids are characterized by extreme variability and large derivative values across multiple signals. Further supporting previous claims.

Table 4.19: Feature values extracted from outlying cycles in Cluster 1 using K-Medoids clustering.

Feature	EP_6_331	CW_7_1207	BH_4_217
Speed skew	0.669	-0.311	0.096
SoC1 σ	1.877	0.925	0.028
Torque skew	1.072	1.415	1.064
DcCurrent diff skew	1.371	-4.984	0.443
Torque mean	2.207	3.359	0.249
Torque $\|\Delta\|$	1.495	4.472	1.263
DcCurrent diff σ	11.597	6.625	1.431
Speed $\|\Delta\|$	166.911	226.183	298.638

Discussion K-Medoids

Overall, K-Medoids produces cluster structures that are consistent with those obtained using HAC and K-Means clustering, while offering increased robustness to extreme observations by utilizing actual data points as cluster centers. This is evident from the selected representative cycles in Table 4.17, which correspond to physically plausible and centrally located cycles rather than averaged behavior. In contrast, outlying cycles identified in Tables 4.18 and 4.19 exhibit pronounced deviations in skewness and derivative-based features without influencing the selection of cluster representatives.

This behavior demonstrates that K-Medoids clustering is less sensitive to atypical and highly varied operational patterns, compared to K-Means as the centroid is an actual data point rather than a mean value that can be affected by extreme values. As a result, K-Medoids is particularly suitable for datasets containing heterogeneous operating conditions where transient events or rare cycles are present.

The identified clusters clearly separate stable and highly dynamic operating regimes, consistent with the patterns observed using HAC and K-Means. Specifically, elevated values of Speed $|\Delta|$, Torque $|\Delta|$, and DcCurrent_diff_σ consistently characterize the high-variability cluster across all three clustering methods. This agreement across different algorithms indicates that the observed structure reflects underlying properties of the engine cycle data rather than artifacts as a result of the algorithm. Consequently, K-Medoids serves as a reliable validation method for the clustering outcomes.

4.3.4 Convolutional Autoencoder Clustering

The results from the CAE, shows the underlying structure of the time series data. The classical clustering algorithms find dominant operating regimes such as high/low load, whilst the auto encoder finds several sub modes within these regimes.

Figure 4.12 presents the clustering validity metrics obtained by applying K-Means to the latent representations learned by the CAE. Unlike the feature based clustering methods, the silhouette score does not exhibit a sharp peak at low values of k but instead shows a gradual improvement up to around $k = 6$. The Calinski–Harabasz index decreases steadily with increasing k, while the Davies–Bouldin index reaches its minimum at higher cluster counts. These results indicate that the latent space learned by the CAE supports a multi cluster structure rather than a simple binary partitioning.

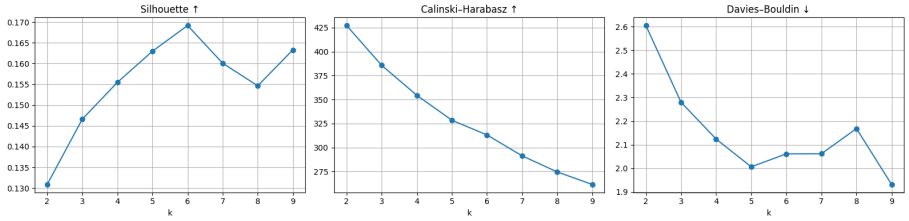


Figure 4.12: The three cluster validity metrics, silhouette score, Calinski-Harabasz and Davies-Bouldin indices for clustering in the CAEs latent space.

Figure 4.13 shows the PCA projection of the latent space for the selected number of clusters $k = 6$. The clusters form partially overlapping but distinct regions, with centroids distributed across the principal axes. Compared to feature based clustering, the separation is less dominated by a single direction of variance, suggesting that the CAE captures a combination of temporal and multivariate signal characteristics.

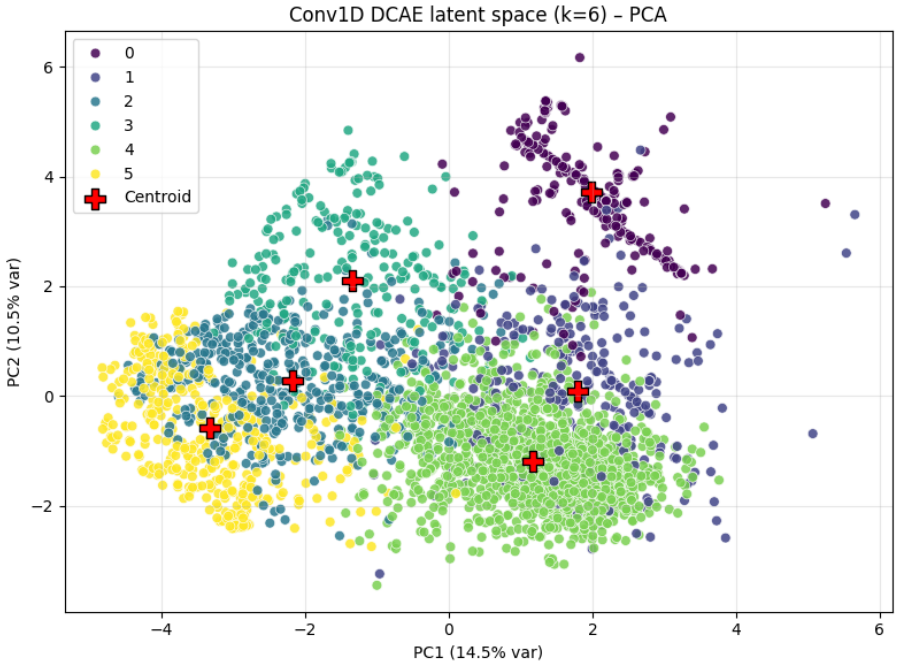


Figure 4.13: PCA visualization of the CAE latent space with $k = 6$.

To further illustrate the structure of the latent space, Figure 4.14 presents a t-SNE embedding of the same representations. The presence of defined groups supports

the use of a higher number of clusters for the CAE based approach.

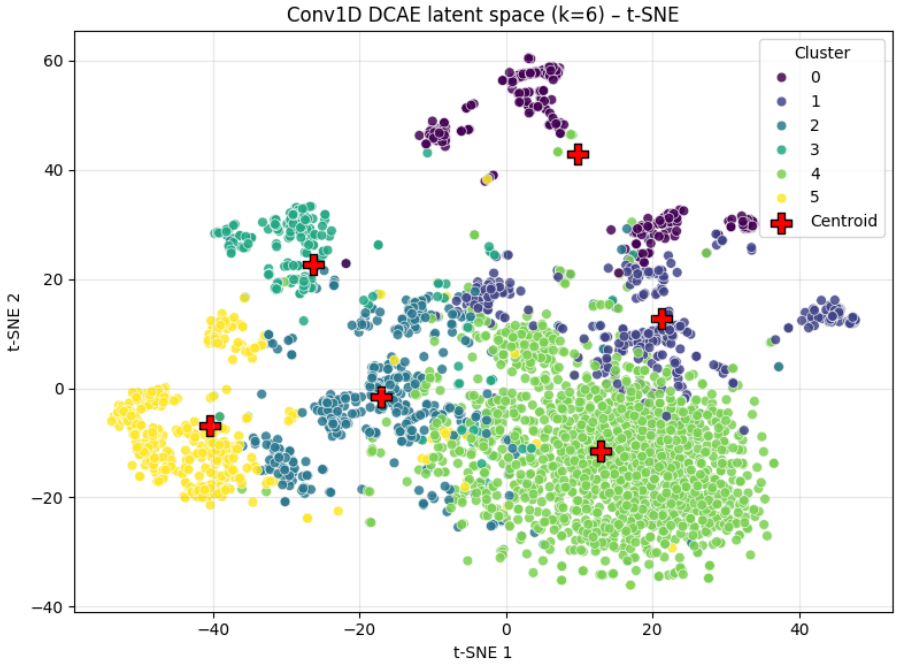


Figure 4.14: t-SNE visualization of the CAE latent space with $k = 6$.

Figure 4.15 summarizes the distribution of mean speed values across the CAE clusters. The clusters again, exhibit distinct operating regimes, ranging from low speed cycles to high speed operation, with clusters that also represent transitional behavior. This demonstrates that, although clustering is performed in latent space, the resulting groups correspond to interpretable differences in driving patterns.

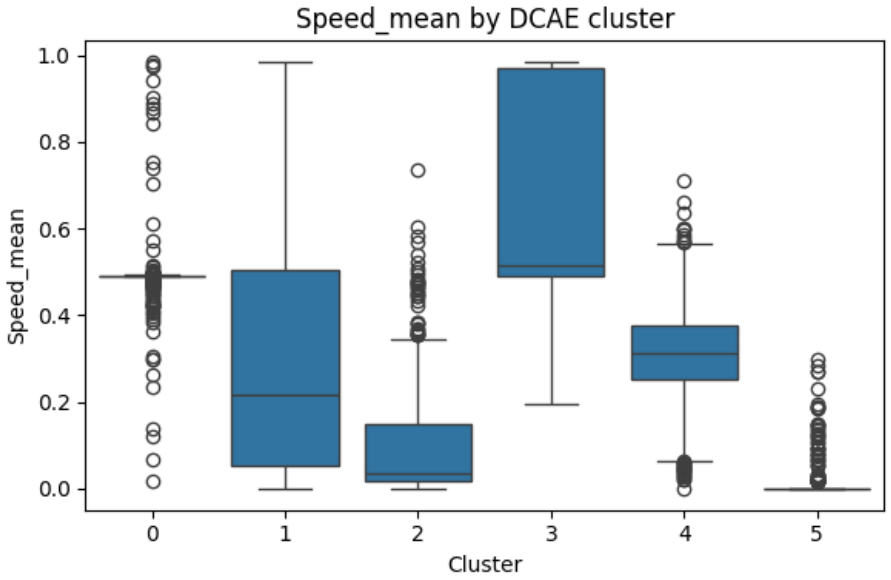


Figure 4.15: Distribution of Speed_mean across CAE based clusters.

Pairwise relationships between selected signal statistics are shown in Figure 4.16. The clusters occupy different regions of the feature space, while preserving the correlations between speed, torque, and current related signals. Unlike feature based clustering, no single statistic dominates the separation, indicating that the CAE integrates temporal dynamics and cross signal interactions when creating its latent representation.

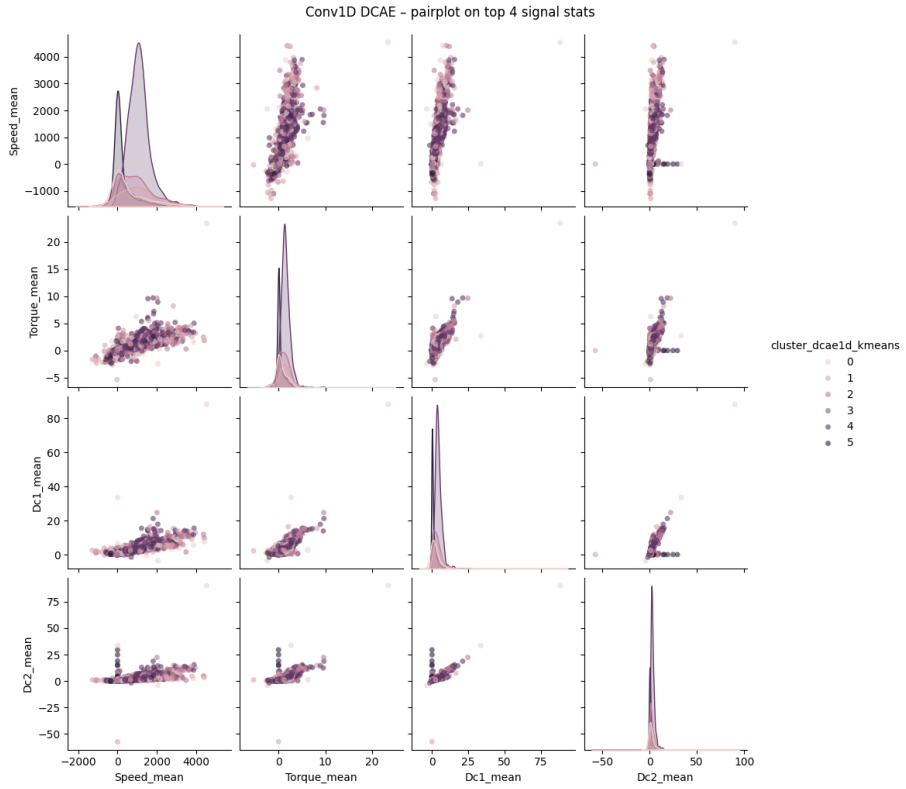


Figure 4.16: Pairwise feature relationships for CAE-based clusters.

Figure 4.17 shows the distribution of CAE clusters across different engines. Each engine contributes cycles to multiple clusters, although cluster 4 is most dominant for CleverWolf_7 and ElegantPanther_6. This suggests that the identified clusters represent actual operating modes that occur across engines rather than engine specific artefacts.

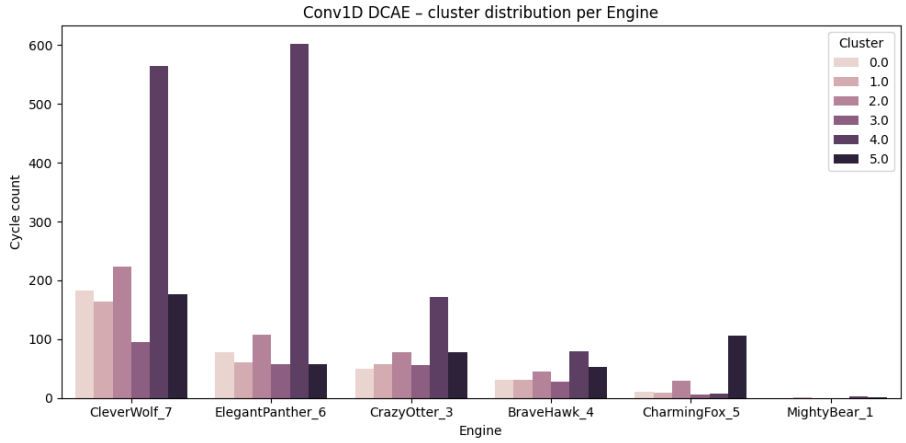


Figure 4.17: Distribution of CAE-based clusters per engine.

Discussion CAE

Overall, the convolutional CAE enables the identification of multiple distinct operating modes by learning a compact latent representation that preserves both global signal occurrences and transient events. Compared to clustering on hand crafted features, the CAE based approach reveals a more nuanced structure in the data, capturing differences in temporal behavior that are not easily separable using statistical analysis alone.

4.4 Drive cycles, Cluster 0

4.4.1 Representative drive cycles for Cluster 0

In order to describe the typical operational characteristics for Cluster 0, the representative drive cycles are analyzed. These cycles are the ones closest to the cluster centroid, in feature space, and therefore reflect the most common operating conditions within the cluster. The representative drive cycles are: drive cycle_429 for CleverWolf_7, drive cycle_604 for CleverWolf_7 and drive cycle_886 for CleverWolf_7. Presented in Table 4.20 is the general information for the drive cycles for the configuration of 8-ton and in Table 4.21 is the general information for the drive cycles for the configuration of 33-ton. One note to add is that *Travel time* and *Traveled distance* may vary between the two vehicle configurations as a result from approximations done during the calculations.

Before presenting the results, we will introduce Figure 4.18. In Figure 4.18 the engine map is shown, i.e. the operational area of the torque and speed of the engine together with the efficiency in the different points. The efficiency is shown with the color map from green with mediate efficiency to yellow with high efficiency. Due to confidentiality the scales, and values for the axis, can not be shown, but in order to relate the torque and speed to low and high values, two black lines are shown. One vertical and one horizontal, representing the reference speed and reference torque, respectively. These correspond to the maximum torque, at maximum power, and the maximum speed, at maximum power. The same reference values are included in all subsequent *Efficiency map* plots as well as in all *Engine torque* and *Engine speed* plots. The two reference lines were also used to compute the reference power level shown in the *Battery behavior* plots, as another black horizontal line.

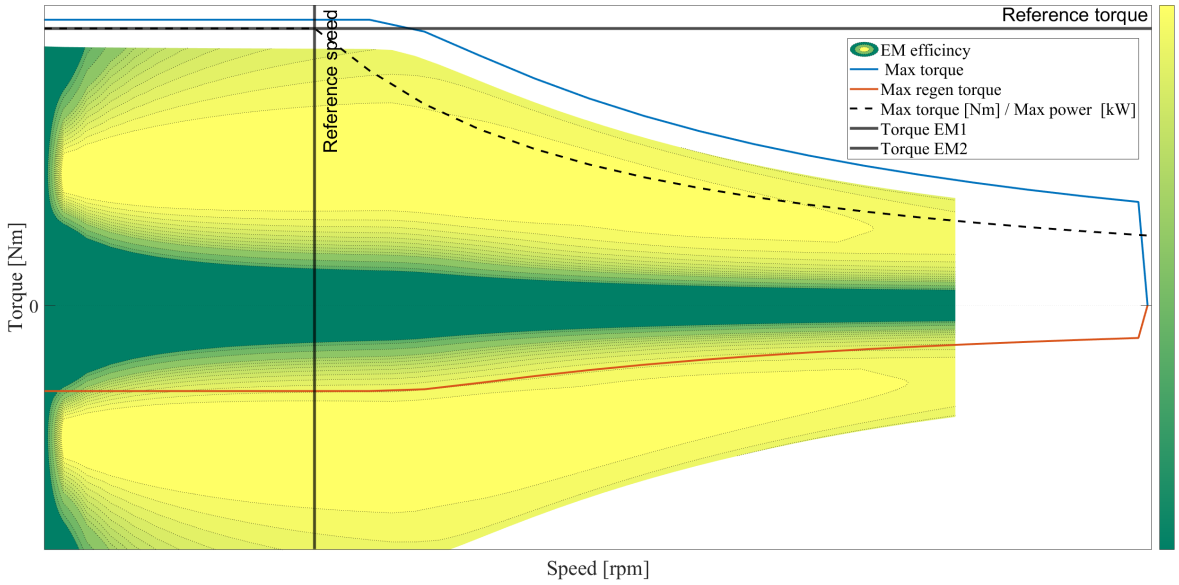


Figure 4.18: An empty engine map used to explain how the reference values for torque and speed has been selected.

The Tables 4.20 and 4.21, highlight low to moderate traveled distances over a narrow time span. The energy consumption is lower for drive cycle_604 when compared to drive cycle_429 and drive cycle_886. Consequently, due to the faster drive cycle and less distance traveled the distance normalized energy consumption is significantly larger for drive cycle_604.

Table 4.20: General information for the 8-ton simulations.

Parameter	Drive cycle_429	Drive cycle_604	Drive cycle_886
Travel time [min]	49	41	47
Traveled distance [km]	3.2601	1.7549	3.6401
Average speed [km/h]	3.9919	2.5942	4.6444
Energy cons [kWh/100km]	145.4	202.2	133.9
Electric cons [kWh]	5.1	3.8	5.2

Table 4.21: General information for the 33-ton simulations.

Parameter	Drive cycle_429	Drive cycle_604	Drive cycle_886
Travel time [min]	48	40	46
Traveled distance [km]	3.2602	1.7552	3.6406
Average speed [km/h]	4.0831	2.6508	4.7355
Energy cons [kWh/100km]	240.7	310.9	241.4
Electric cons [kWh]	8.4	5.9	9.4

Electromagnetic torque and speed

Figures 4.19-4.21 shows the torque and speed for both motors for the three representative drive cycles for the 8-ton configuration. Figures 4.22-4.24 shows the same data but for the 33-ton configuration. As can be seen in the figures the drive cycles has moderate traction phases, with similar amplitudes. Both vehicle configurations, Figures 4.19-4.21 and Figures 4.22-4.24, exhibit clear operating events separated by idle periods. During the operational phases the torque and speed increase, and engine one continuously reaches higher rotational speed magnitudes. This is because engine one has several gear ratios, while engine two has a fixed ratio, resulting in engine one being able to operate at different gears. When engine one operates at lower gears higher rotational speeds are reached, and when the engines operate at the same gears they reach equal rotational speeds. The 8-ton configuration highlight torque fluctuations for both positive and negative values, indicating both propulsion and regenerative braking operation.

The 33-ton configuration, see Figures 4.22-4.24, operates very similarly to the 8-ton configuration, but exhibits more frequent variation in torque. This is a direct result of the increased vehicle mass. Consequently, a shift in torque magnitude and duration occur, due to the increased inertia. Larger and more frequent magnitudes of negative torque can also be seen, reflecting the increased braking needed due to larger inertia. Both vehicle configurations exhibit dominating start-stop operation rather than continuous propulsion, characteristic for drive cycles in Cluster 0. Both engines are active, indicating a shared workload during operation, regardless of mass.

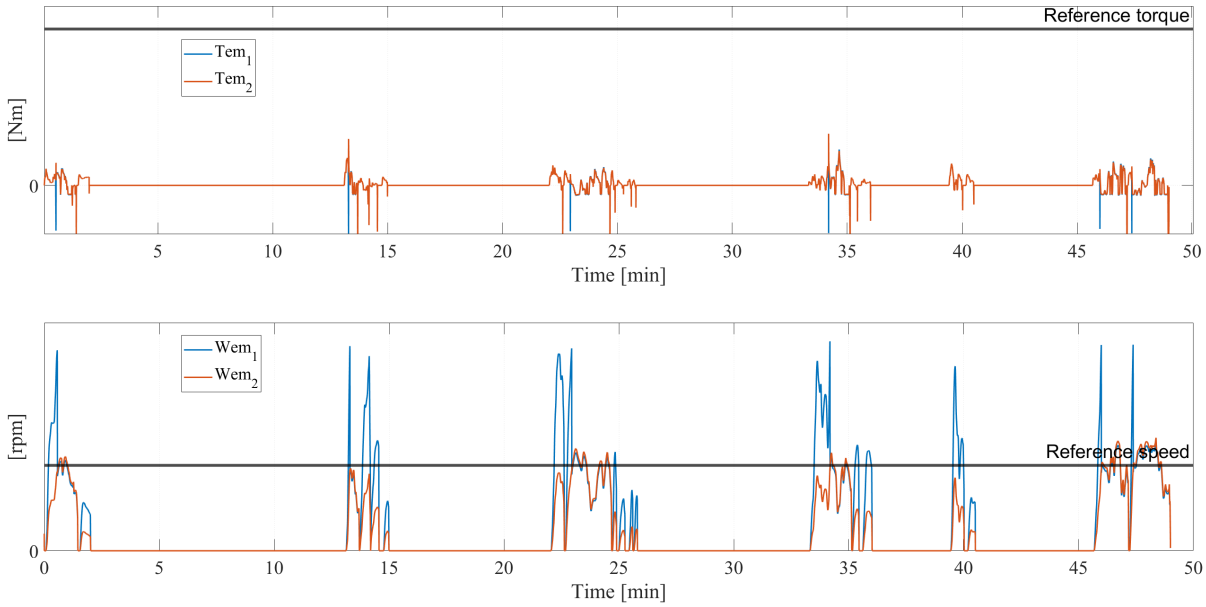


Figure 4.19: The electromagnetic torque and speed for the representative drive cycle_429, ElegantPanther_6, in the 8-ton configuration.

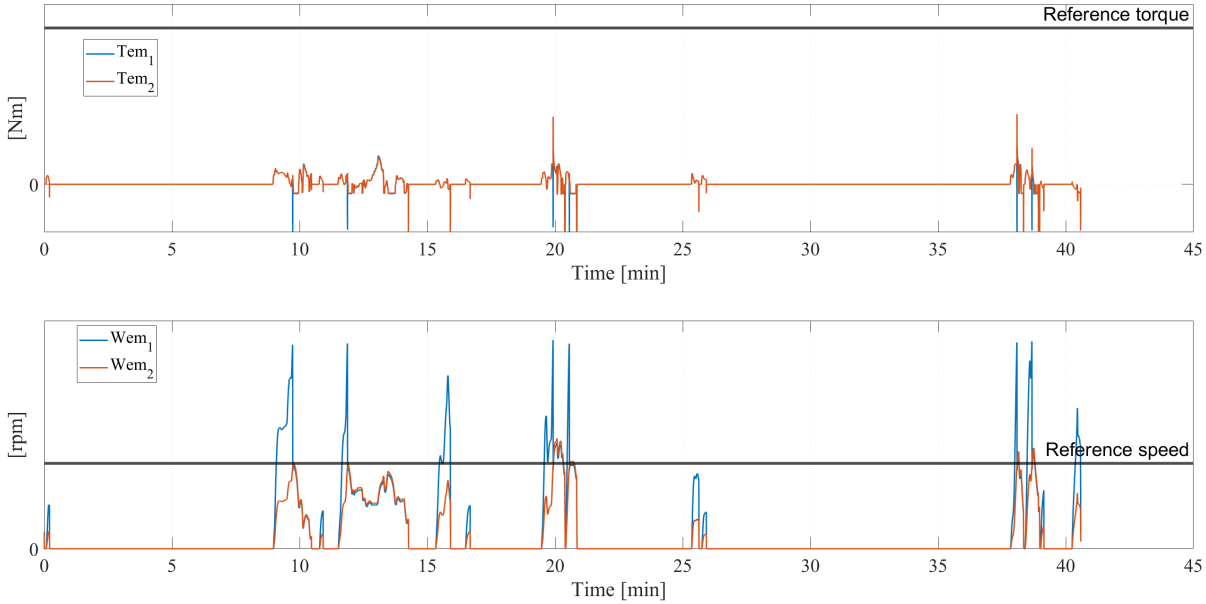


Figure 4.20: The electromagnetic torque and speed for the representative drive cycle_604, ElegantPanther_6, in the 8-ton configuration.

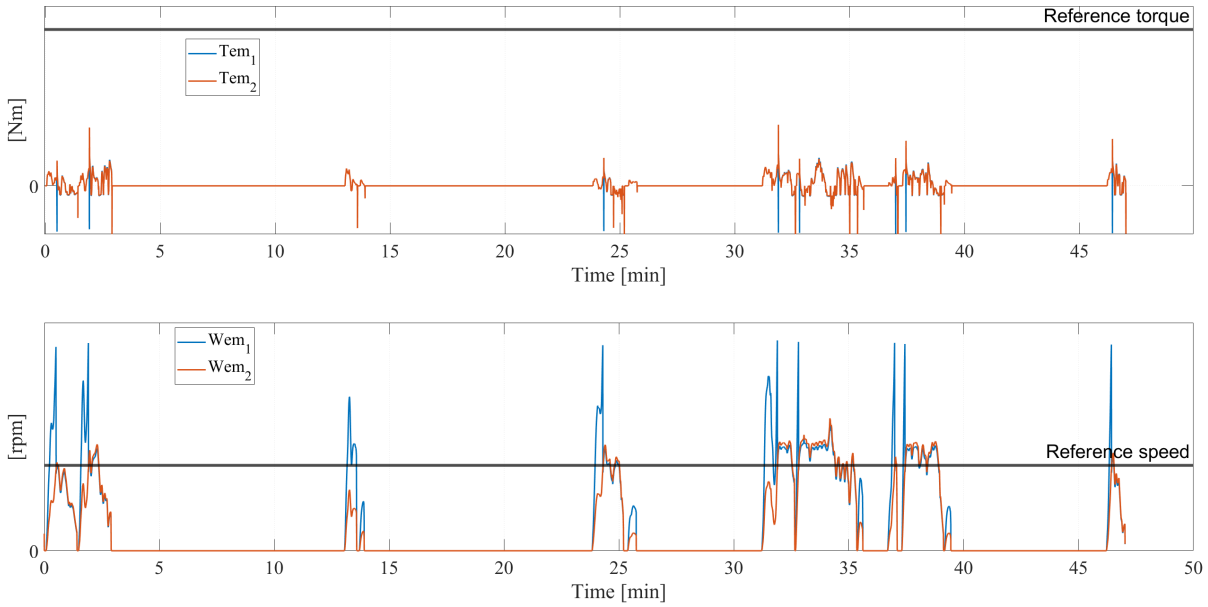


Figure 4.21: The electromagnetic torque and speed for the representative drive cycle_886, ElegantPanther_6, in the 8-ton configuration.

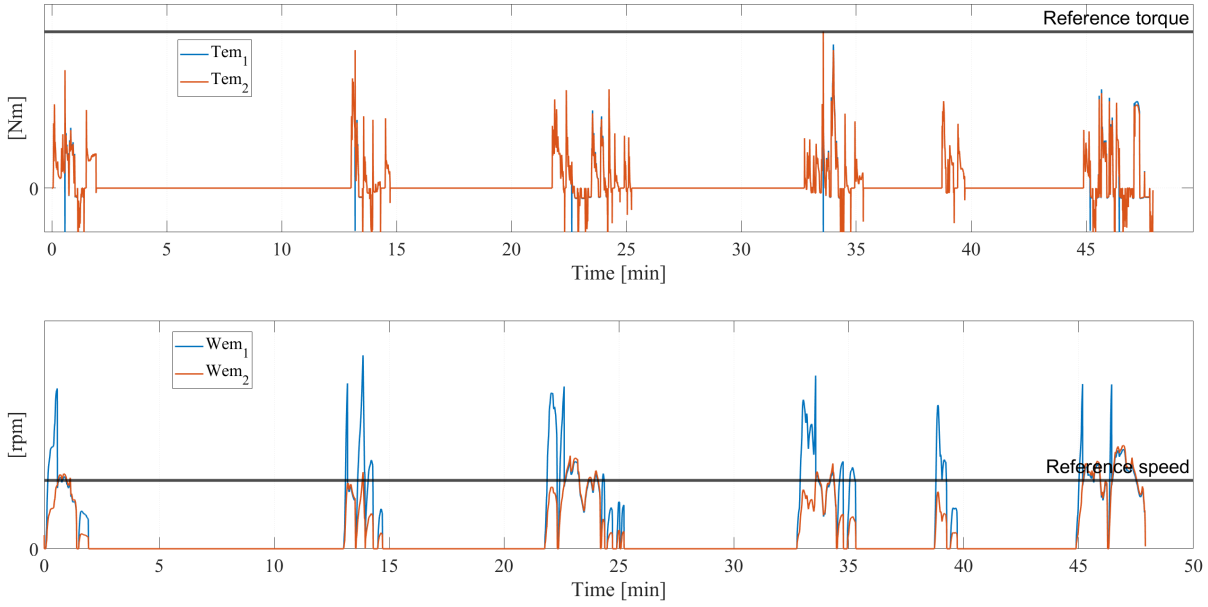


Figure 4.22: The electromagnetic torque and speed for the representative drive cycle_429, ElegantPanther_6, in the 33-ton configuration.

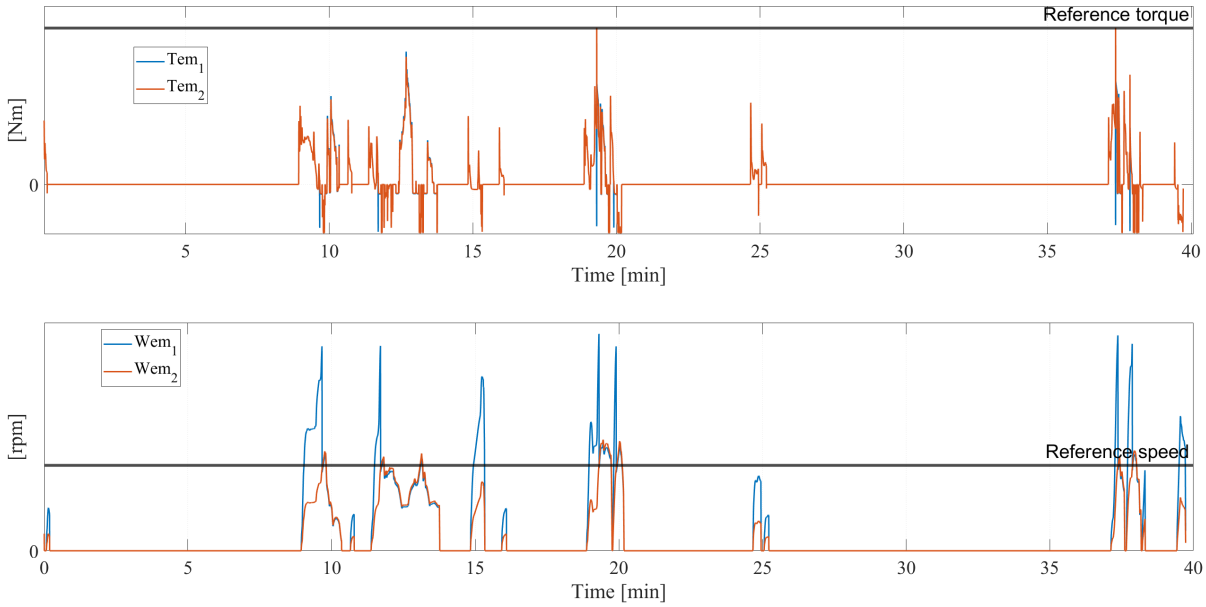


Figure 4.23: The electromagnetic torque and speed for the representative drive cycle_604, ElegantPanther_6, in the 33-ton configuration.

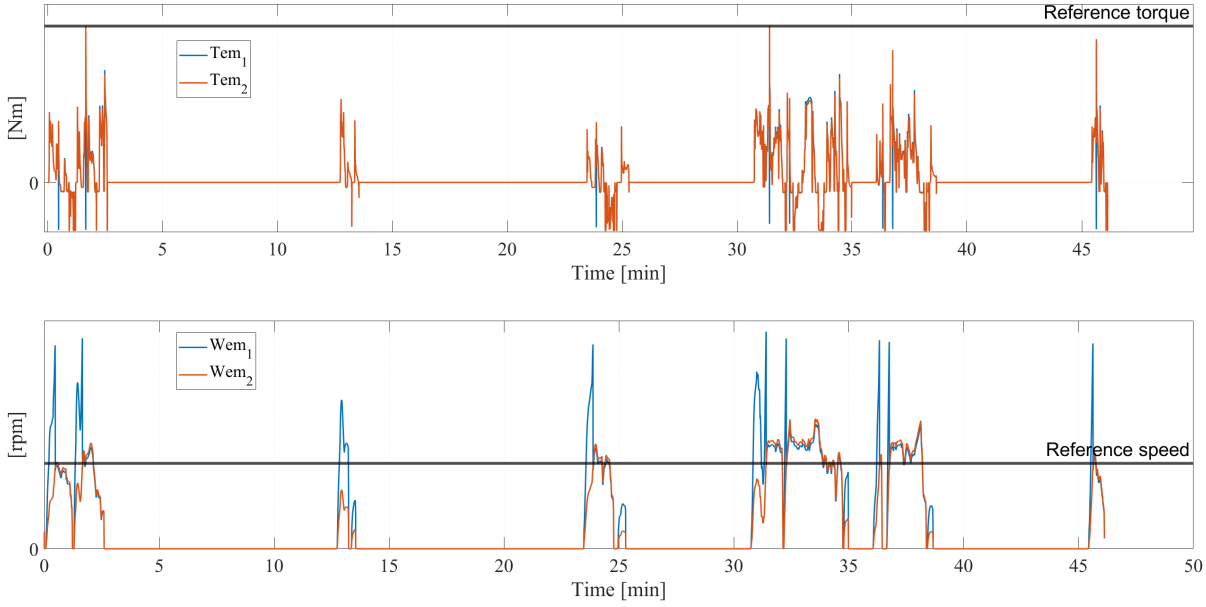


Figure 4.24: The electromagnetic torque and speed for the representative drive cycle_886, ElegantPanther_6, in the 33-ton configuration.

Battery Behavior

Starting of with the 8-ton configuration, Figures 4.25-4.27 exhibits how the battery acts during the drive cycle operations. For the 8-ton configuration the C_*rate* profile is characterized by low to moderate values with short-lived peaks occurring throughout the drive cycle. These peaks correspond to the traction demand and the regenerative braking. The periods with low C_*rate* indicate idle- or low load conditions. For the 33-ton configuration, see Figures 4.28-4.30, an increase in C_*rate* magnitude peaks and battery power is presented. This reflects the increased power demand consistent with a higher mass of the vehicle.

The battery state of charge (SOC) exhibits a near linear decrease over the drive cycle duration. This reflects consistent energy consumption without sudden and/or abrupt events, consisting of the moderate power demand observed from the C_{rate} profile. The battery power fluctuations occur around zero [kW], indicating balanced switching between discharge- and regenerative operation. The battery power is well within the charging power limit throughout the whole drive cycle as well. As for the SOC curve for the 33-ton configuration, despite the increased load, it remains near linear without sudden drops. The battery power, mainly the regenerated power, extends into larger magnitudes during regenerative braking events, indicating increased battery stress following the increased mass but even so the battery still operates well within the limit range.

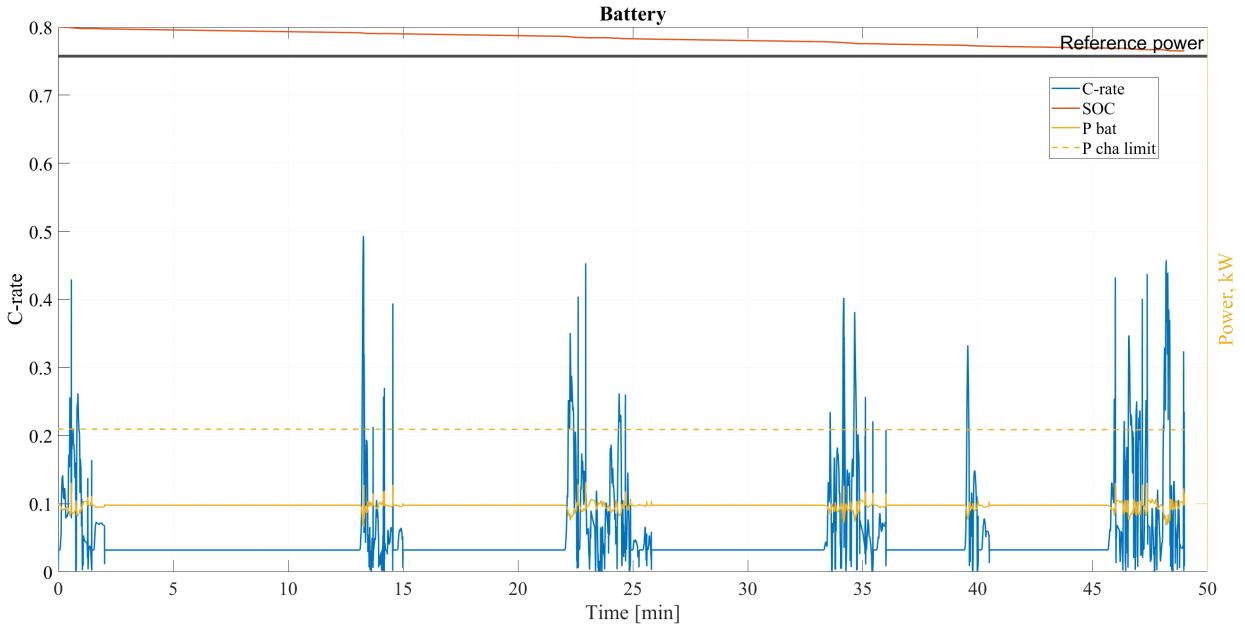


Figure 4.25: The battery behavior detected for the representative drive cycle_429, ElegantPanther_6, in the 8-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

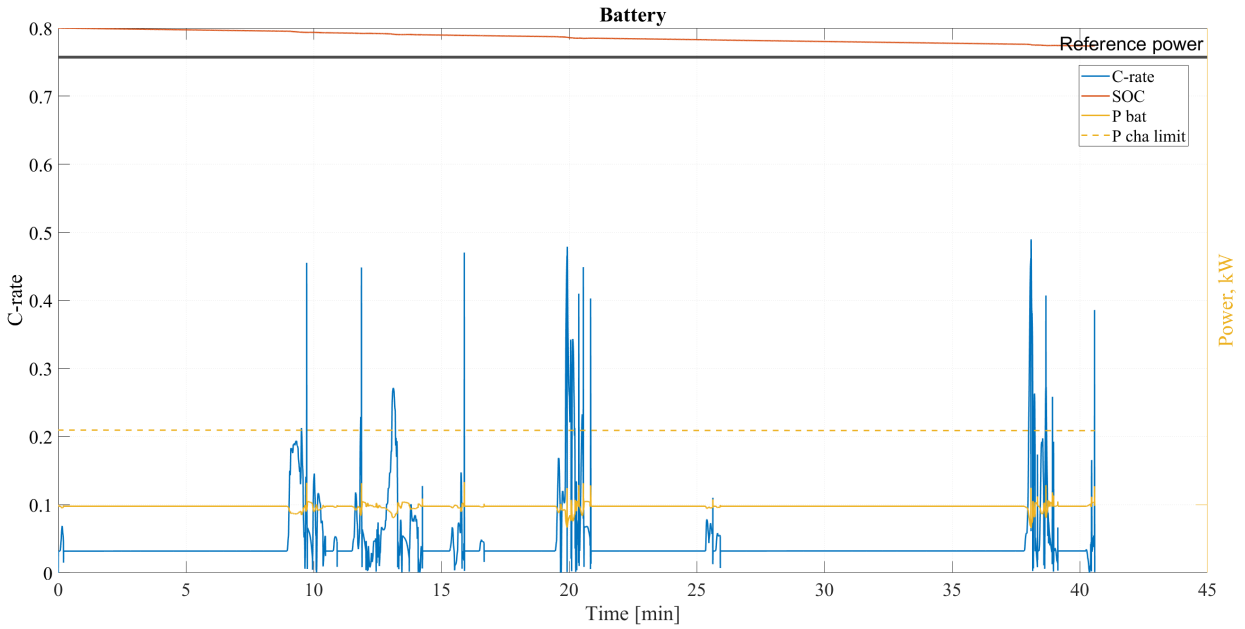


Figure 4.26: The battery behavior detected for the representative drive cycle_604, ElegantPanther_6, in the 8-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

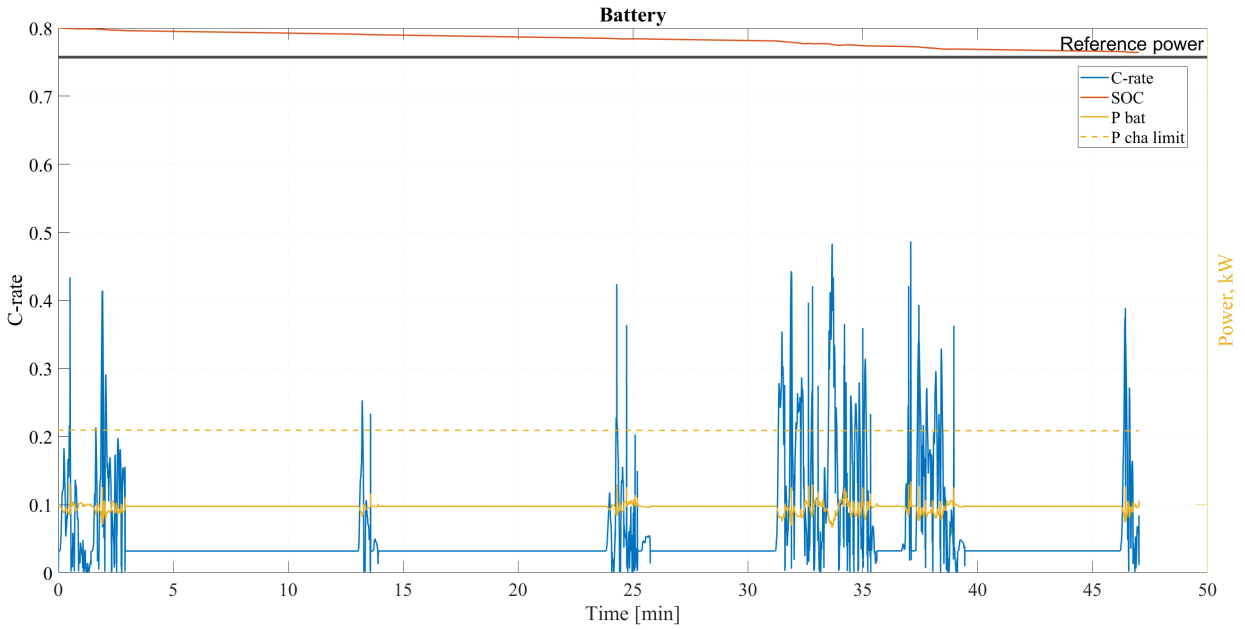


Figure 4.27: The battery behavior detected for the representative drive cycle_886, ElegantPanther_6, in the 8-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

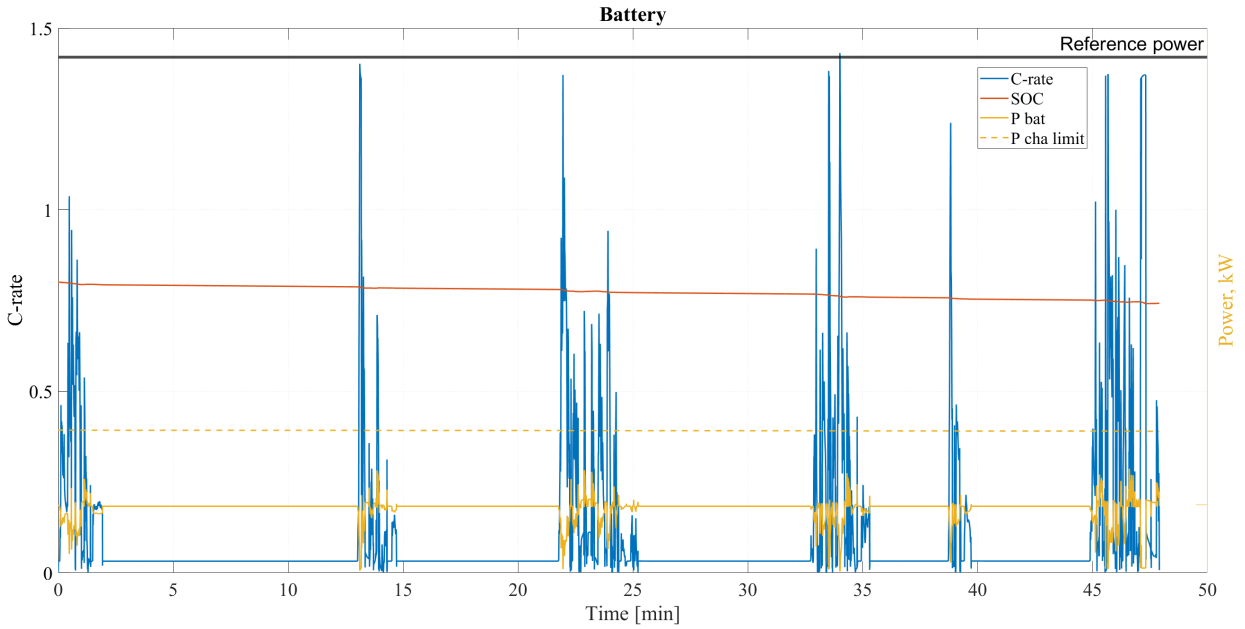


Figure 4.28: The battery behavior detected for the representative drive cycle_429, ElegantPanther_6, in the 33-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

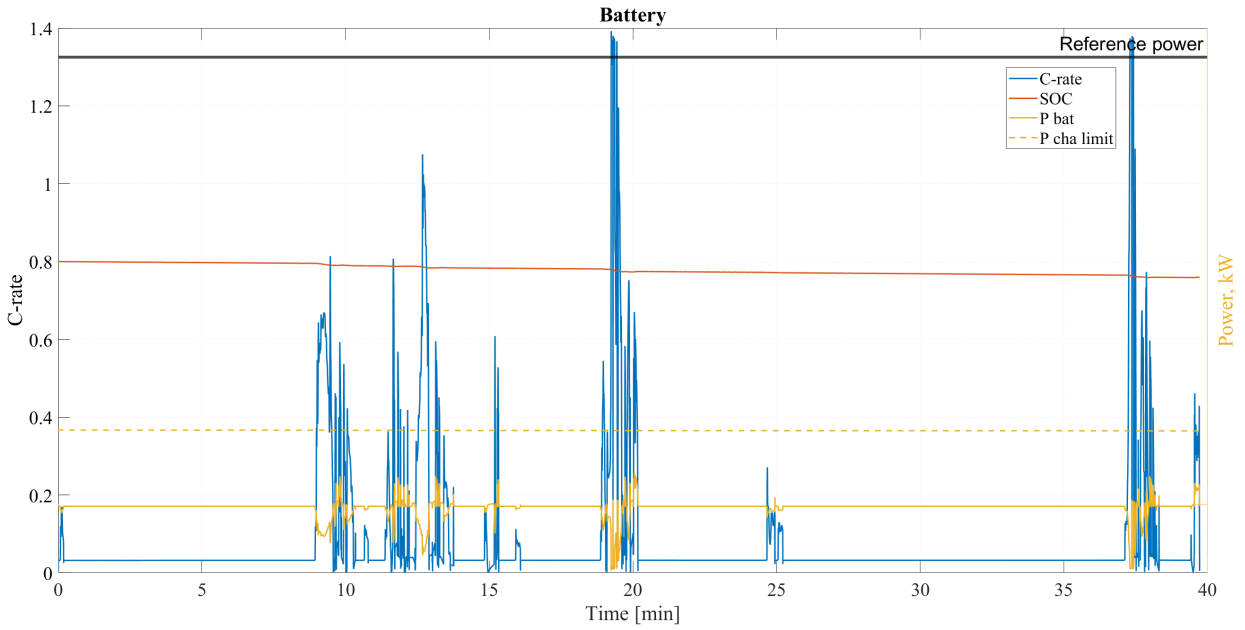


Figure 4.29: The battery behavior detected for the representative drive cycle_604, ElegantPanther_6, in the 33-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

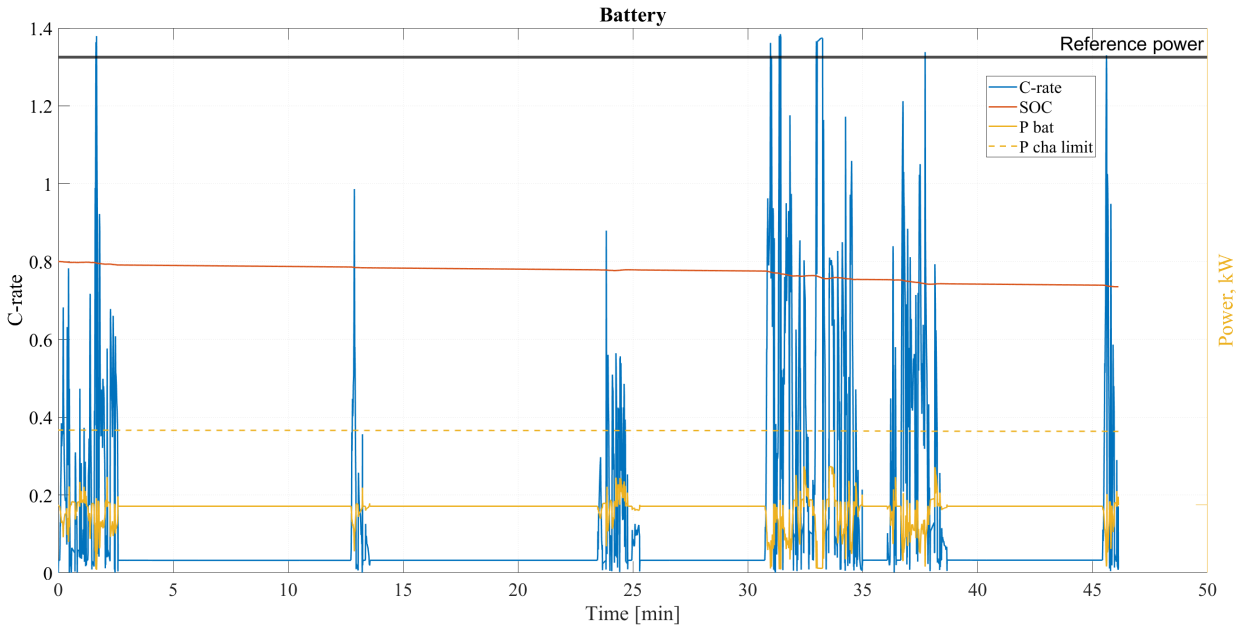


Figure 4.30: The battery behavior detected for the representative drive cycle_886, ElegantPanther_6, in the 33-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

The energy metrics in Tables 4.22 and 4.23, exhibit little variation in SOC during the drive cycles. The variation is slightly increased for the heavier vehicle configuration but still small. For both vehicle configurations, the SOC reduction is relatively small, reflecting the short- and "gentle" operation across the drive cycles, as well as the moderate energy demand. The captured energy is present but very small when compared to the total energy output. The Energy Storage System (ESS) efficiency is consequently above 98% through out all drive cycles, regardless of vehicle configuration.

Table 4.22: Energy storage data for the 8-ton simulations.

Parameter	Drive cycle_429	Drive cycle_604	Drive cycle_886
Integrated current [Ah]	7.16	5.36	7.36
Total Energy Output [kWh]	5.5	4.1	5.7
Total Energy captured [kWh]	0.38	0.25	0.45
ESS efficiency [%]	99.8	99.8	99.7
Energy throughput [kWh/h]	7.15	6.36	7.81
End SOC [%]	76.51	77.39	76.41
SOC difference [%-points]	3.49	2.61	3.59
Energy charge station [kWh]	0	0	0

Table 4.23: Energy storage data for the 33-ton simulations.

Parameter	Drive cycle_429	Drive cycle_604	Drive cycle_886
Integrated current [Ah]	11.86	8.25	13.27
Total Energy Output [kWh]	9.6	6.7	10.9
Total Energy captured [kWh]	1.26	0.87	1.67
ESS efficiency [%]	99.1	99.2	98.9
Energy throughput [kWh/h]	13.58	11.39	16.41
End SOC [%]	74.22	75.98	73.53
SOC difference [%-points]	5.78	4.02	6.47
Energy charge station [kWh]	0	0	0

Engine behavior

For the 8-ton vehicle configuration, the electric machine (engine), operating points are mainly concentrated at low to moderate torque levels and the speed ranges from low to moderate levels. The majority of operating points lie within the region of moderate efficiency, indicating good engine utilization for the representative drive cycles of Cluster 0. This is further supported by Table 4.24, which exhibits the mean-value engine efficiency during the drive cycles.

Table 4.24: Electric machine efficiency data for the 8-ton simulations.

Parameter	Drive cycle_429	Drive cycle_604	Cycle 886
EM 1, mean efficiency [%]	85.25	86.26	86.69
EM 2, mean efficiency [%]	85.90	86.74	86.66

Figures 4.31-4.33 and Figures 4.34-4.36 also highlights that the operating points are within the traction and regenerative operating limits. There are larger values detected for the maximum regenerative torque at low speeds for the 8-ton configuration. The grouped operating points reflects stable low conditions characteristic for Cluster 0, once again supporting the previous drive cycle results, see Chapter 4.3.

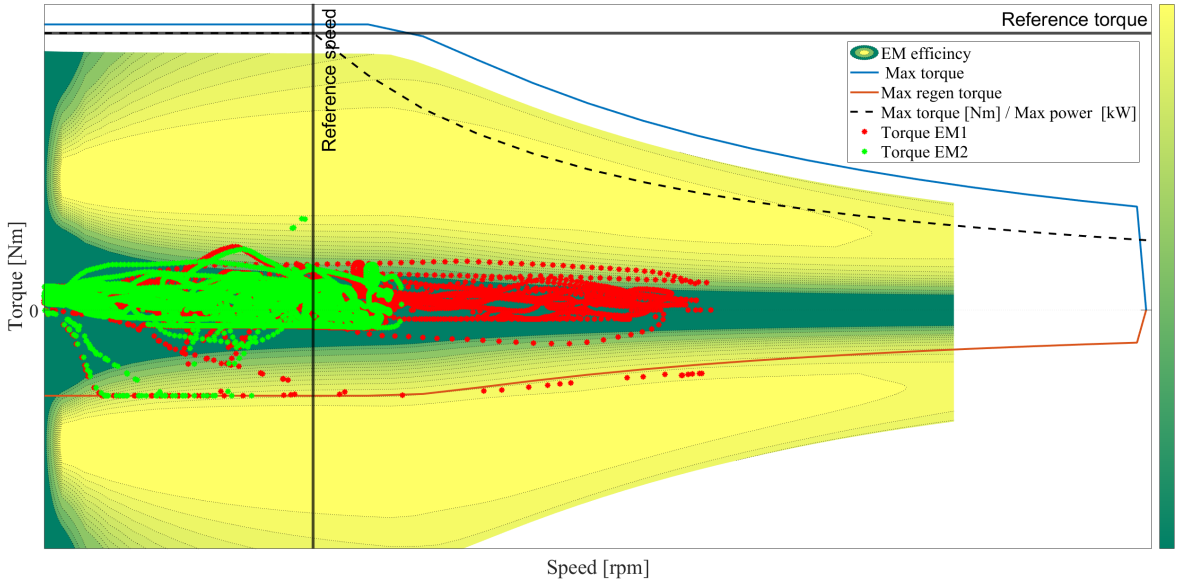


Figure 4.31: Engine behavior for the representative drive cycle_429, ElegantPanther_6, in the 8-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

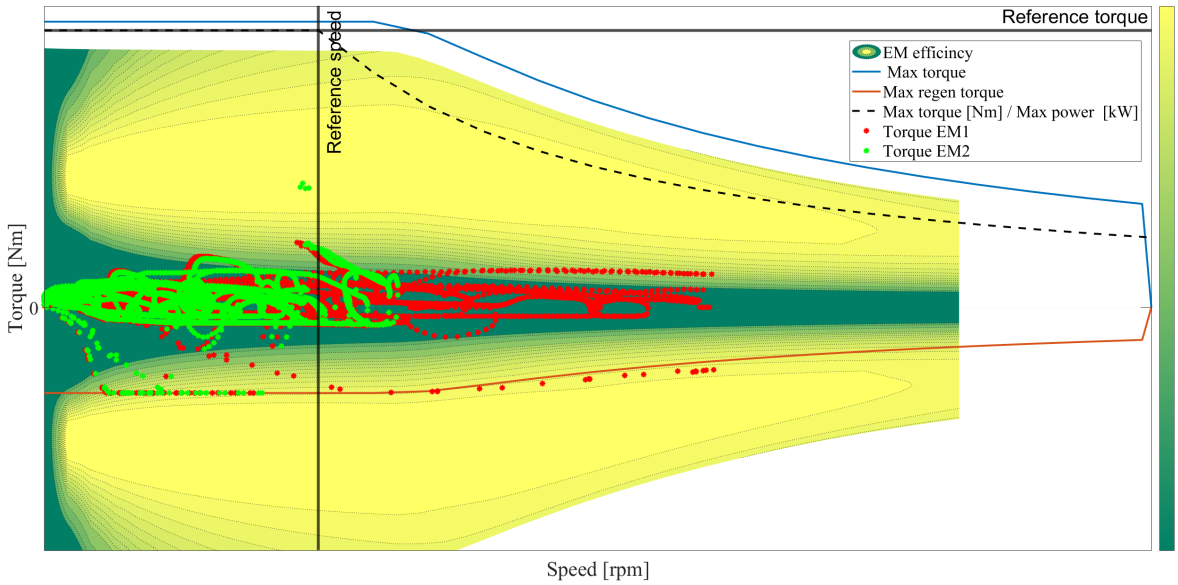


Figure 4.32: Engine behavior for the representative drive cycle_604, ElegantPanther_6, in the 8-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

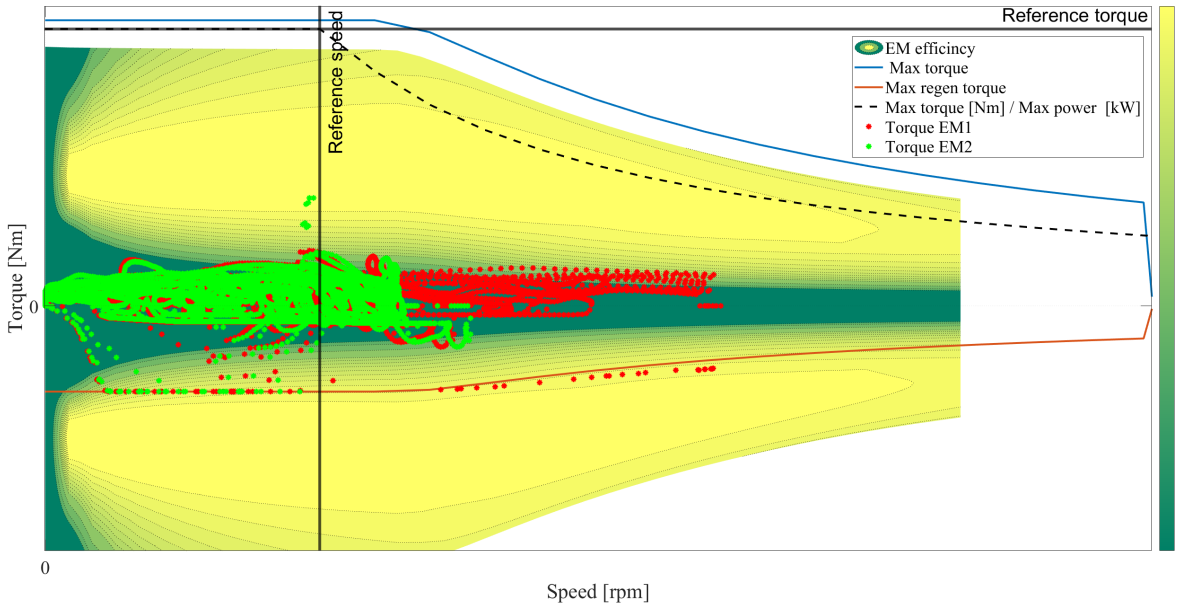


Figure 4.33: Engine behavior for the representative drive cycle_886, ElegantPanther_6, in the 8-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

For the 33-ton configuration, Figures 4.34-4.36, the increased mass is reflected in the higher torque levels. A wider speed span is detected as well. The distribution of operating points shifts upwards in the torque plane and the speed profile remains similar to the that of the 8-ton configuration. Despite the increased mass, the operating points are group at higher efficiency values, which also is shown in Table 4.25, which exhibit the mean-value engine efficiency during the drive cycles. As for the regenerative braking events, they are spread across a wider region in the torque-speed plane but still within the limit.

Table 4.25: Electric machine efficiency data for the 33-ton simulations.

Parameter	**Drive cycle_429**	**Drive cycle_604**	**Cycle 886**
EM 1, mean efficiency [%]	90.96	91.49	92.05
EM 2, mean efficiency [%]	90.80	90.93	91.97

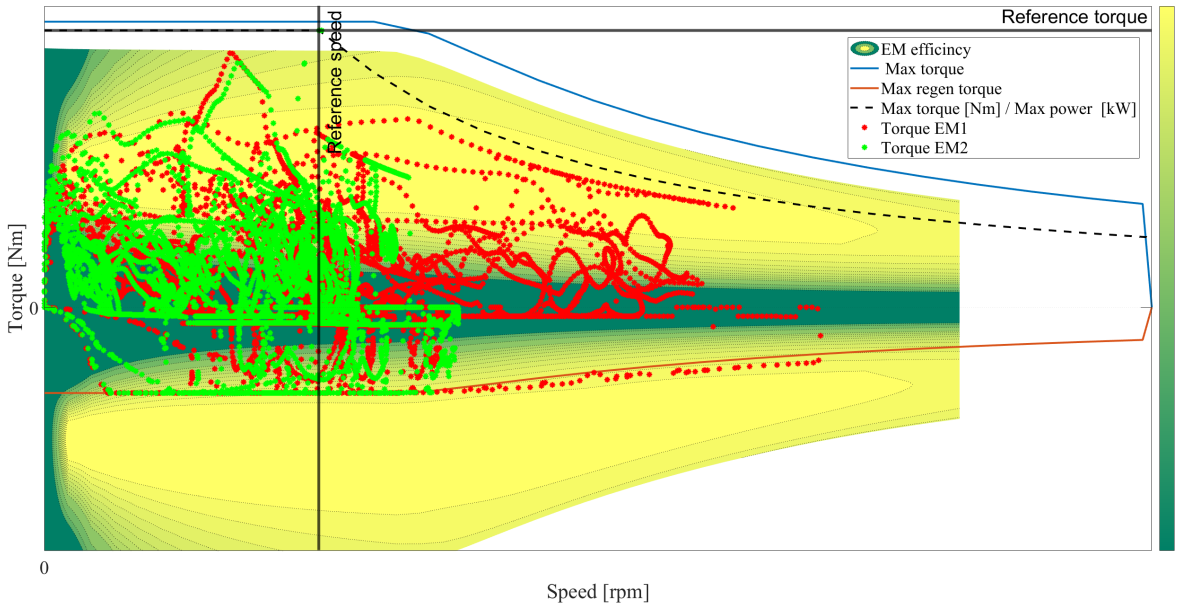


Figure 4.34: Engine behavior for the representative drive cycle_429, ElegantPanther_6, in the 33-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

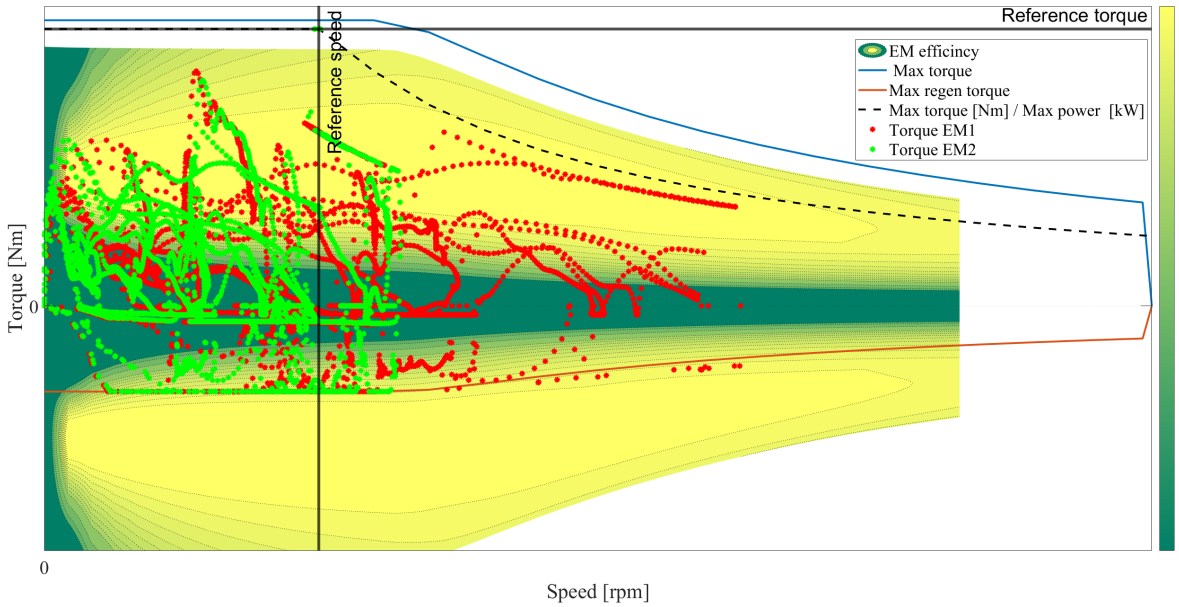


Figure 4.35: Engine behavior for the representative drive cycle_604, ElegantPanther_6, in the 33-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

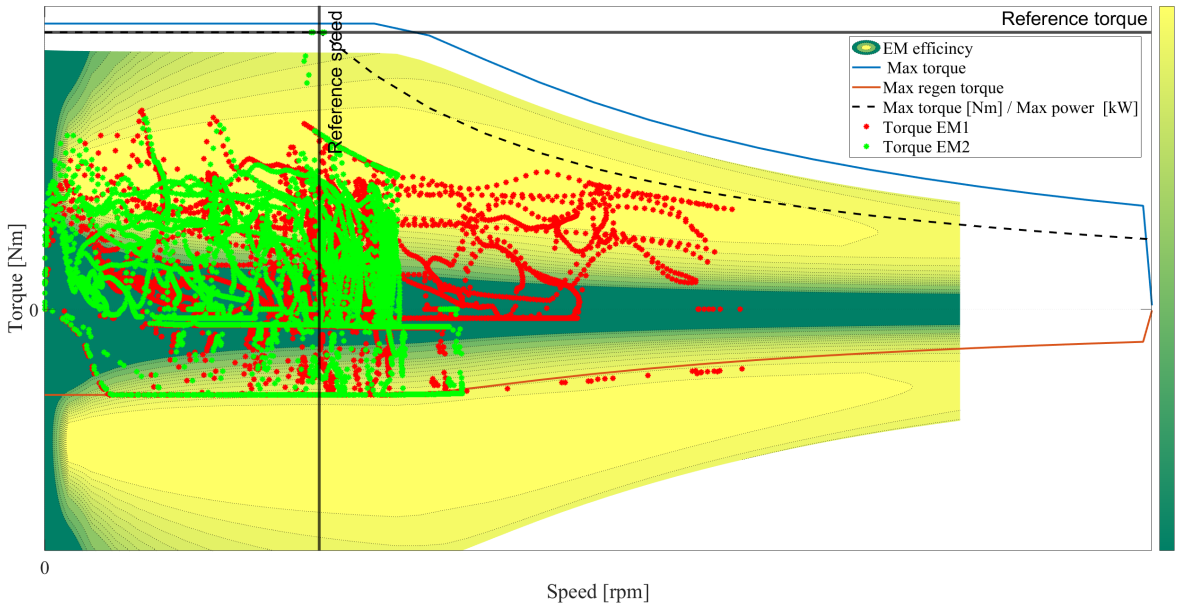


Figure 4.36: Engine behavior for the representative drive cycle_886, ElegantPanther_6, in the 33-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

When comparing the two configurations the most noticeable difference is the shift in operating torque. Both configurations mainly operate within high efficiency regions, emphasizing that Cluster 0 operates efficiently. The similarity between the two configurations indicates that the increased mass mainly influences the instantaneous torque demand and not the main drive cycle operation. The overall engine behavior observed confirms that Cluster 0 represents stable and efficient system operation across both vehicle configurations.

4.4.2 Outlying drive cycles for Cluster 0

In order to describe outlying operational characteristics for Cluster 0, the outlying drive cycles were examined as well. These drive cycles are the ones furthest away from the cluster centroid, in feature space, and therefore reflect the most atypical operating conditions within the cluster. The outlying drive cycles are: drive cycle_547 for CrazyOtter_3, drive cycle_369 for BraveHawk_4 and drive cycle_55 for CharmingFox_5.

One note to add for the general information presented in Table 4.26 and Table 4.27 is that *Travel time* and *Traveled distance* may vary between the two vehicle configurations as a result from approximations done during the calculations. Table 4.26 and Table 4.27, exhibit a clear deviation in drive cycle duration when compared to the representative drive cycles, see Table 4.20 and Table 4.21. The traveled

distance is significantly shorter as well. This results in the distance normalized energy consumption being significantly higher than for those of the representative drive cycles.

Table 4.26: General information for the 8-ton simulations.

Parameter	Drive cycle_196	Drive cycle_906	Drive cycle_1722
Travel time [min]	5	2	38
Traveled distance [km]	0.10044	0.46436	2.5583
Average speed [km/h]	1.2417	14.1549	4.0736
Energy cons [kWh/100km]	389.9	106.4	140.0
Electric cons [kWh]	0.4	0.5	3.9

Table 4.27: General information for the 33-ton simulations.

Parameter	Drive cycle_196	Drive cycle_906	Drive cycle_1722
Travel time [min]	5	2	37
Traveled distance [km]	0.10048	0.46454	2.5582
Average speed [km/h]	1.2639	15.2727	4.1486
Energy cons [kWh/100km]	506.6	331.3	209.3
Electric cons [kWh]	0.5	1.7	5.8

Electromagnetic torque and speed

The electromagnetic torque and speed profiles for the 8-ton vehicle configuration highlight a clear sequence of traction and idle phases. During the positive traction intervals, both engine one and engine two exhibit positive torque followed by a corresponding increase in rotational speed. Once again engine one reaches higher speed levels due to more gear ratios than engine two, as previously stated in subsection 4.4.1. Figures 4.37-4.39 exhibit moderate torque- and speed fluctuations indicating stable torque control during operation.

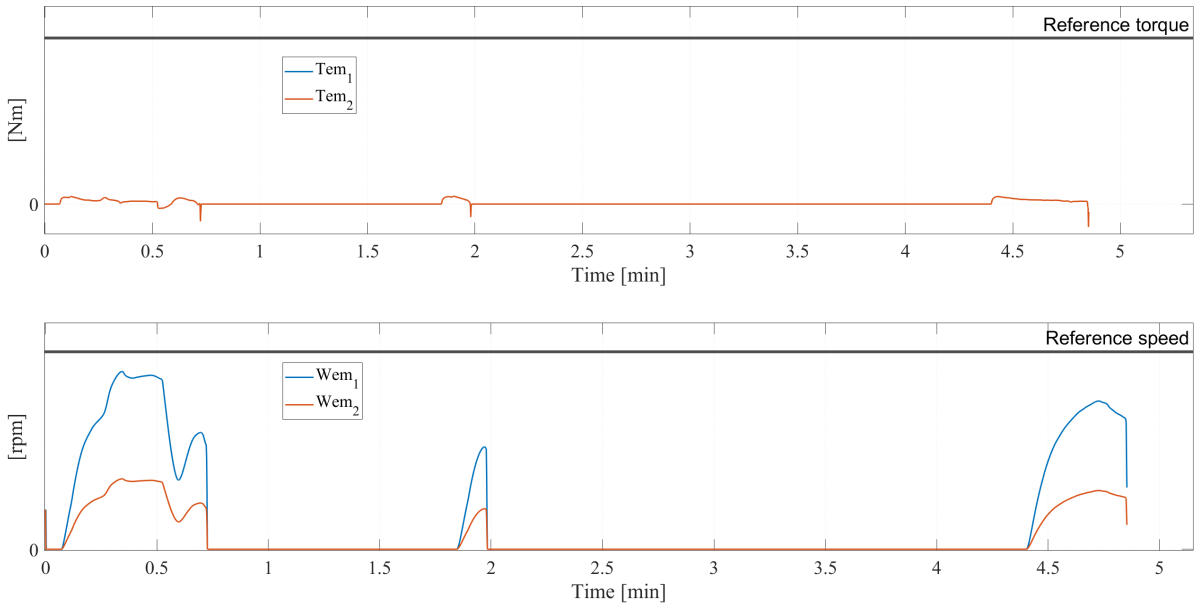


Figure 4.37: Electromagnetic torque and speed for outlying drive cycle_196, BraveHawk_4, in the 8-ton configuration for Cluster 0.

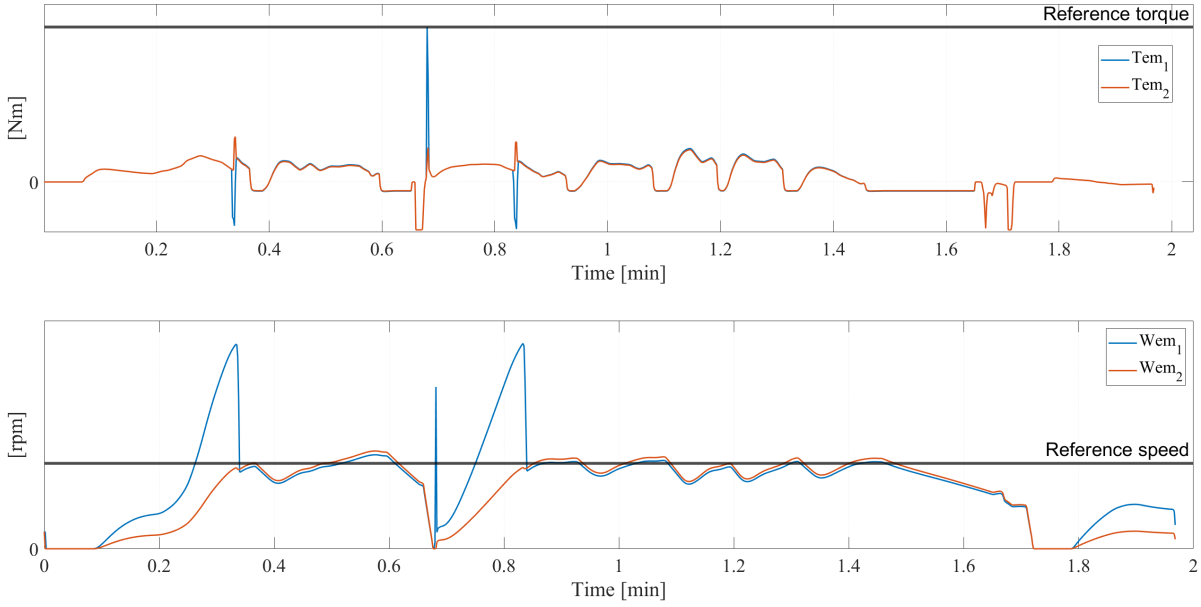


Figure 4.38: Electromagnetic torque and speed for outlying drive cycle_906, ElegantPanther_6, in the 8-ton configuration for Cluster 0.

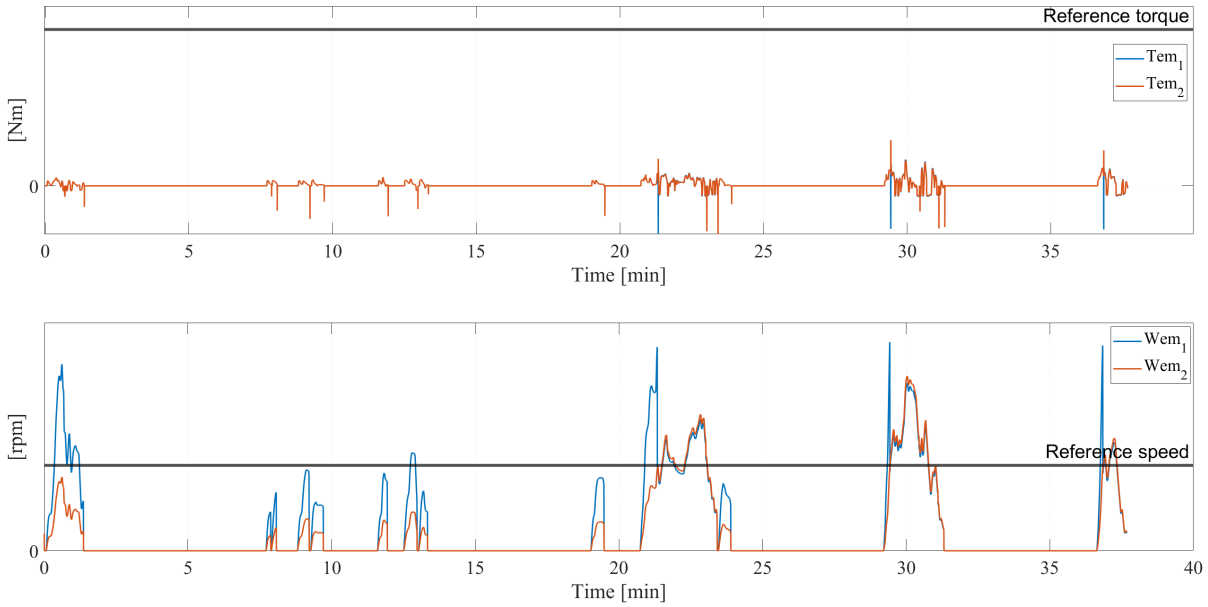


Figure 4.39: Electromagnetic torque and speed for outlying drive cycle_1722, CleverWolf_7, in the 8-ton configuration for Cluster 0.

For the 33-ton configuration, see Figures 4.40-4.42, the torque displays enhanced torque oscillations during operation. The increased torque fluctuations reflects the added inertia and emphasizes dynamic control in order to maintain propulsion during operation. Following the increased inertia is also more clearly pronounced braking- and regenerative events for the 33-ton configuration. Lastly, even though the 33-ton configuration has longer torque operating events, when compared to the 8-ton configuration, long idle phases still occur.

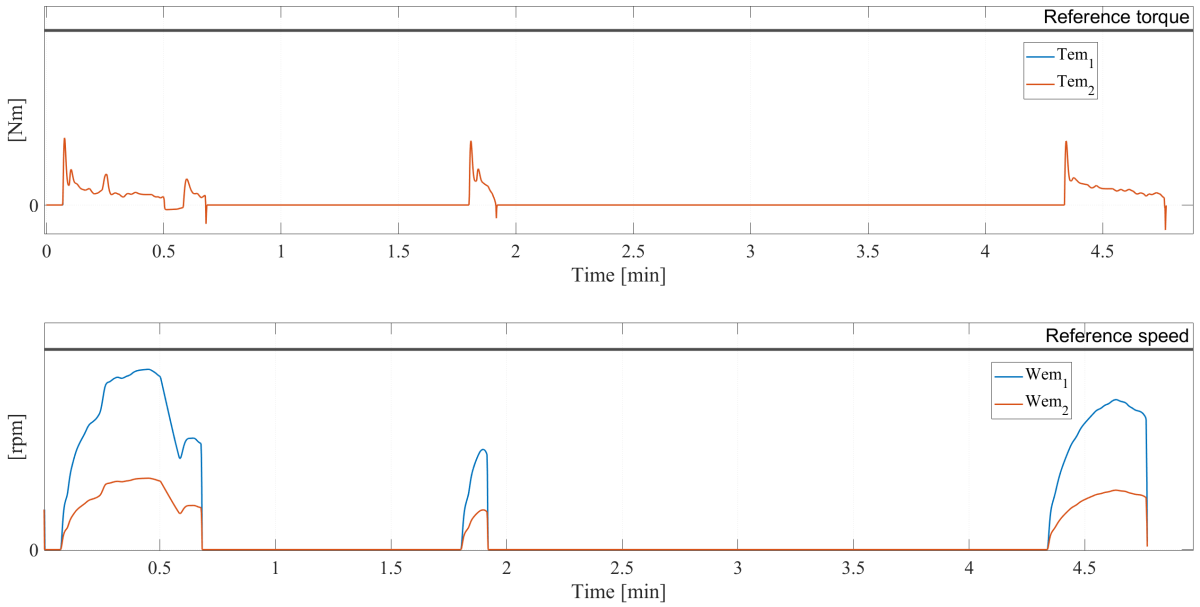


Figure 4.40: Electromagnetic torque and speed for outlying drive cycle_196, BraveHawk_4, in the 33-ton configuration for Cluster 0.

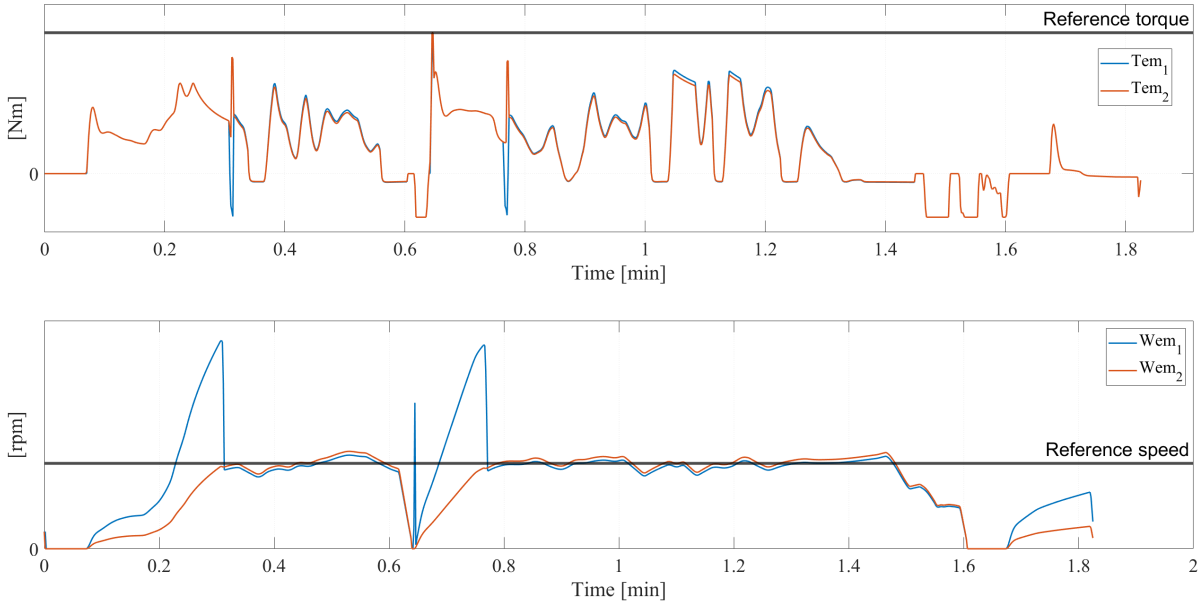


Figure 4.41: Electromagnetic torque and speed for outlying drive cycle_906, ElegantPanther_6, in the 33-ton configuration for Cluster 0.

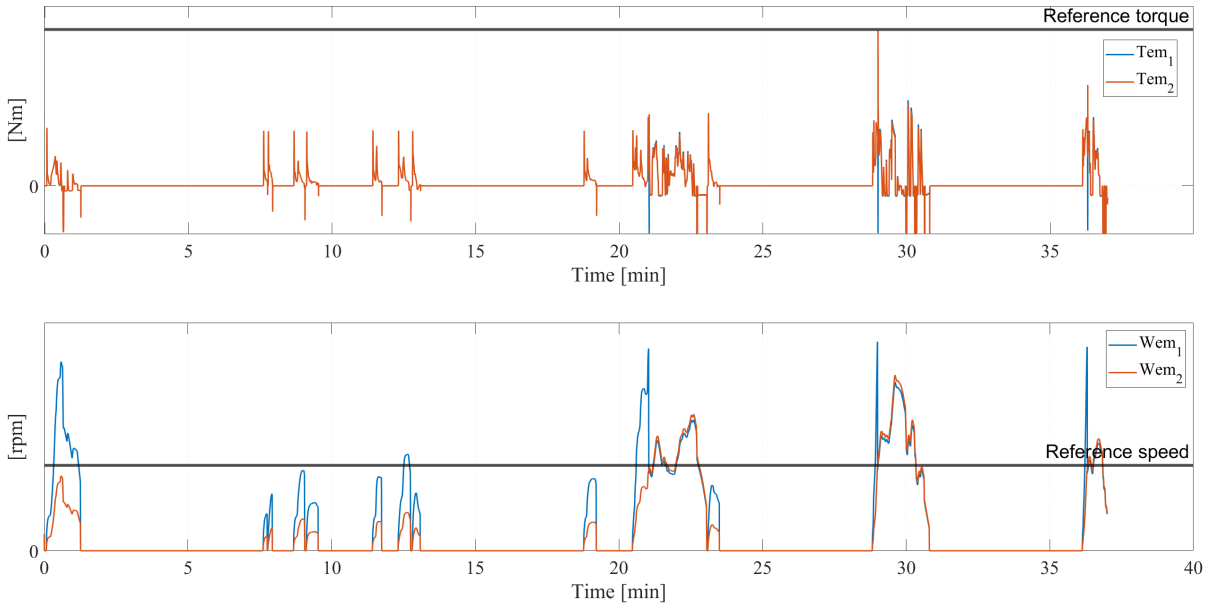


Figure 4.42: Electromagnetic torque and speed for outlying drive cycle_1722, CleverWolf_7, in the 33-ton configuration for Cluster 0.

Battery behavior

Both vehicle configurations for the outlying drive cycles for Cluster 0 highlight battery behavior characterized by longer operation with reduced transient loading when compared to the representative drive cycles. This corresponds with fewer bursts and peaks of the C_{rate}. Figures 4.43-4.45, presents the 8-ton configuration and Figures 4.46-4.48, presents the 33-ton configuration.

From the ESS perspective, the outlying drive cycles exhibit significantly lower values for both total energy output and total captured energy (regenerated energy). The main reason for this is the short operating duration for drive cycles _196 and _906. On the other hand, drive cycle_1722 exhibits disproportionally high energy throughput compared to the traveled distance. Despite this, the total ESS efficiency remains high for all drive cycles throughout both vehicle configurations which is seen in Table 4.28 and Table 4.29.

Table 4.28: Energy storage data for the 8-ton simulations.

Parameter	Drive cycle_196	Drive cycle_906	Drive cycle_1722
Integrated current [Ah]	0.59	0.75	5.42
Total Energy Output [kWh]	0.4	0.6	4.1
Total Energy captured [kWh]	0.00	0.06	0.21
ESS efficiency [%]	99.9	99.6	99.8
Energy throughput [kWh/h]	5.23	20.15	6.78
End SOC [%]	79.71	79.64	77.36
SOC difference [%-points]	0.29	0.36	2.64
Energy charge station [kWh]	0	0	0

Table 4.29: Energy storage data for the 33-ton simulations.

Parameter	Drive cycle_196	Drive cycle_906	Drive cycle_1722
Integrated current [Ah]	0.77	2.32	8.09
Total Energy Output [kWh]	0.5	1.8	6.3
Total Energy captured [kWh]	0.00	0.13	0.61
ESS efficiency [%]	99.9	98.5	99.2
Energy throughput [kWh/h]	6.91	62.27	11.23
End SOC [%]	79.63	78.87	76.06
SOC difference [%-points]	0.37	1.13	3.94
Energy charge station [kWh]	0	0	0

As previously mentioned the outlying drive cycles in Cluster 0 exhibit battery behavior consistent with a more consistent driving pattern when compared to the representative drive cycles. The transient events are fewer with extended intervals in between. There is also significantly lower SOC levels reached. The decreased variability, distinguishes the outliers from the representative drive cycles, where battery operation is more evenly distributed over time.

Both vehicle configurations display similar behavior, differences arise in the magnitude and duration of the SOC events. The 33-ton configuration exhibits higher instantaneous battery power and peaks of C_{rate}, reflecting the increased energy demand resulting from higher vehicle mass. The 8-ton configuration exhibits less frequent and longer-lived spikes, indicating a more even demand in the battery profile. Despite these differences the battery operation for both configurations remains well within the defined power and charging limits. There are no continuous operation near the battery limit levels and peak charging, as well as discharging events occur only briefly.

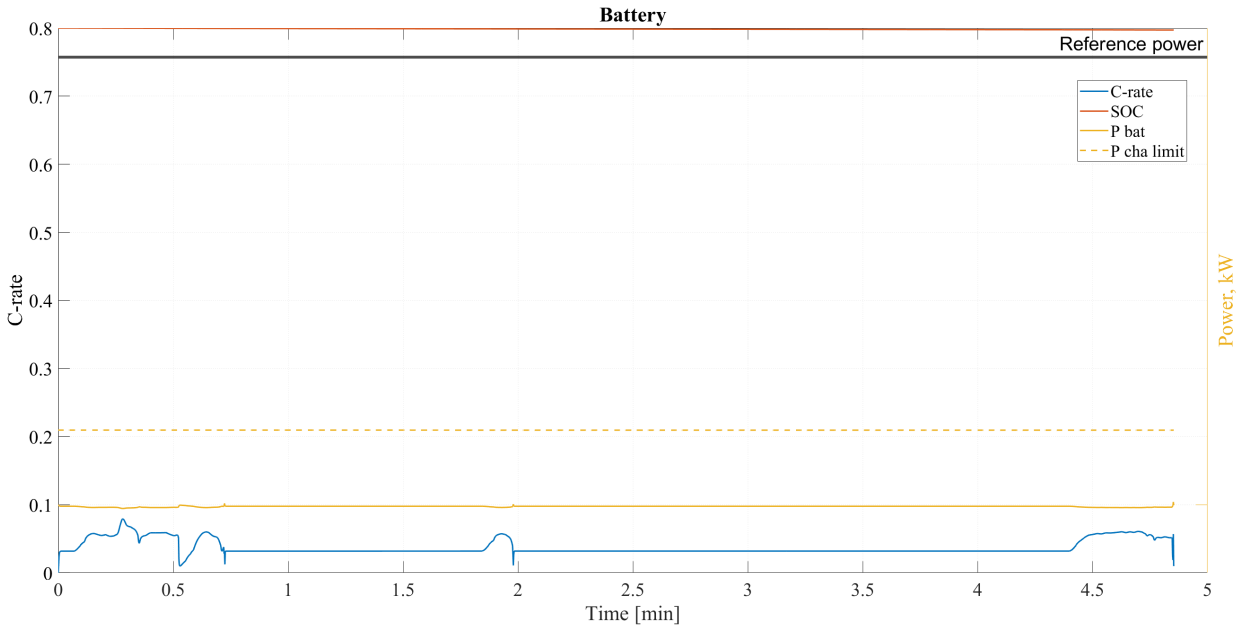


Figure 4.43: The battery behavior detected for the outlying drive cycle_196, BraveHawk_4, in the 8-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

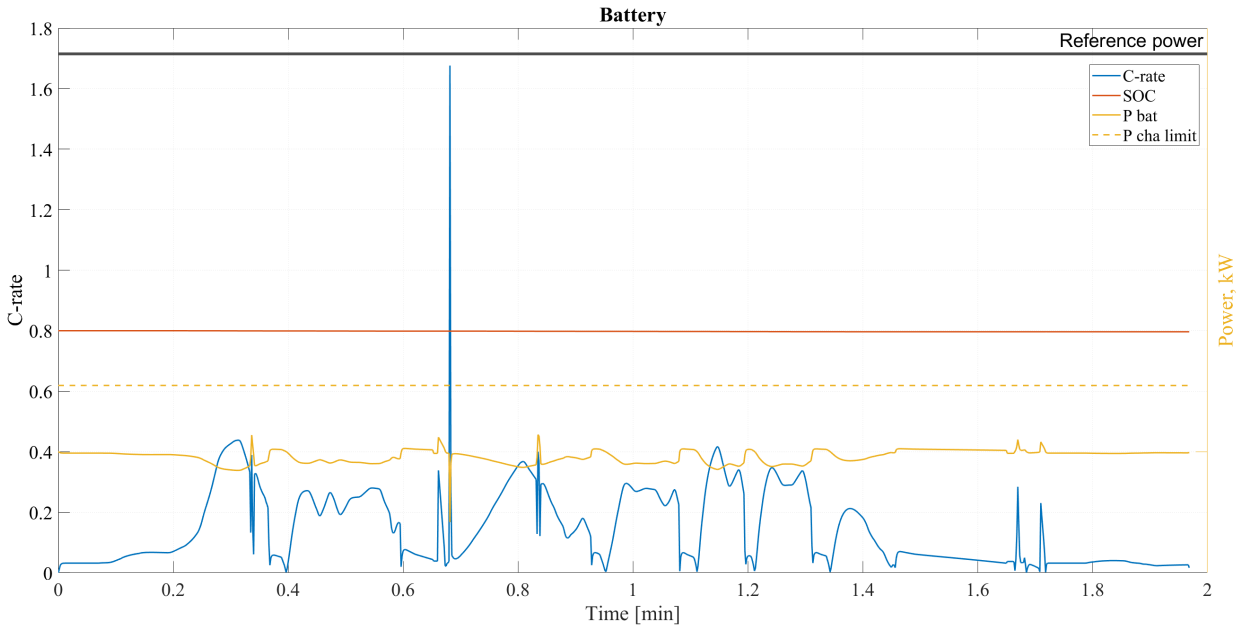


Figure 4.44: The battery behavior detected for the outlying drive cycle_906, ElegantPanther_6, in the 8-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

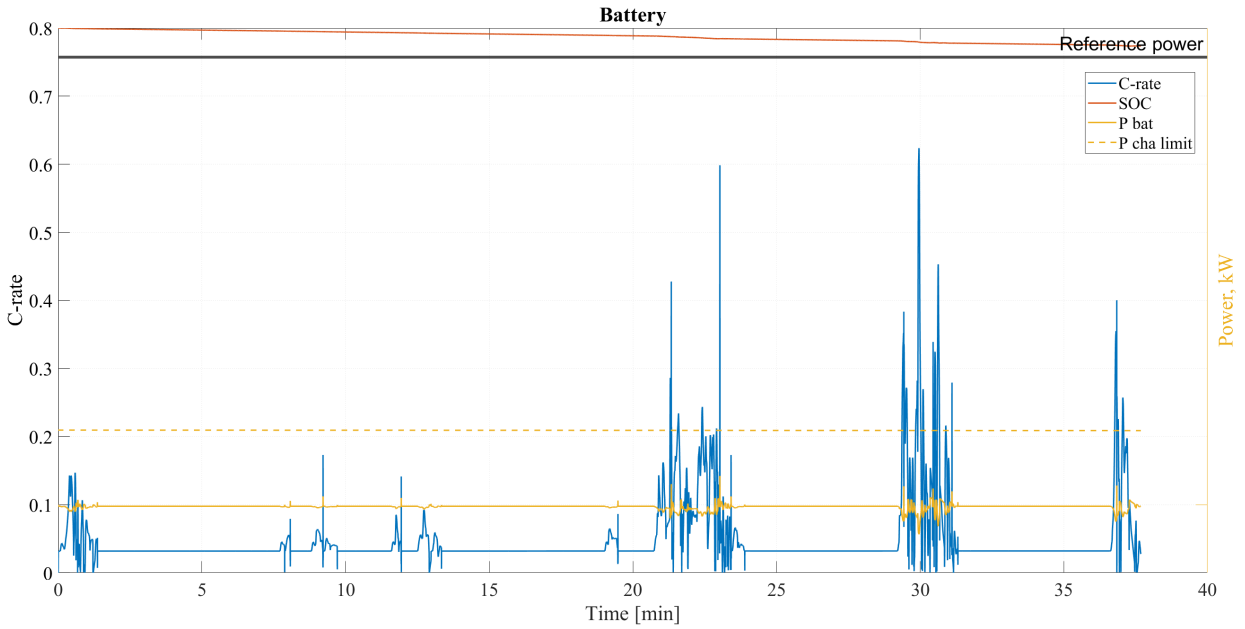


Figure 4.45: The battery behavior detected for the outlying drive cycle_1722, CleverWolf_7, in the 8-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

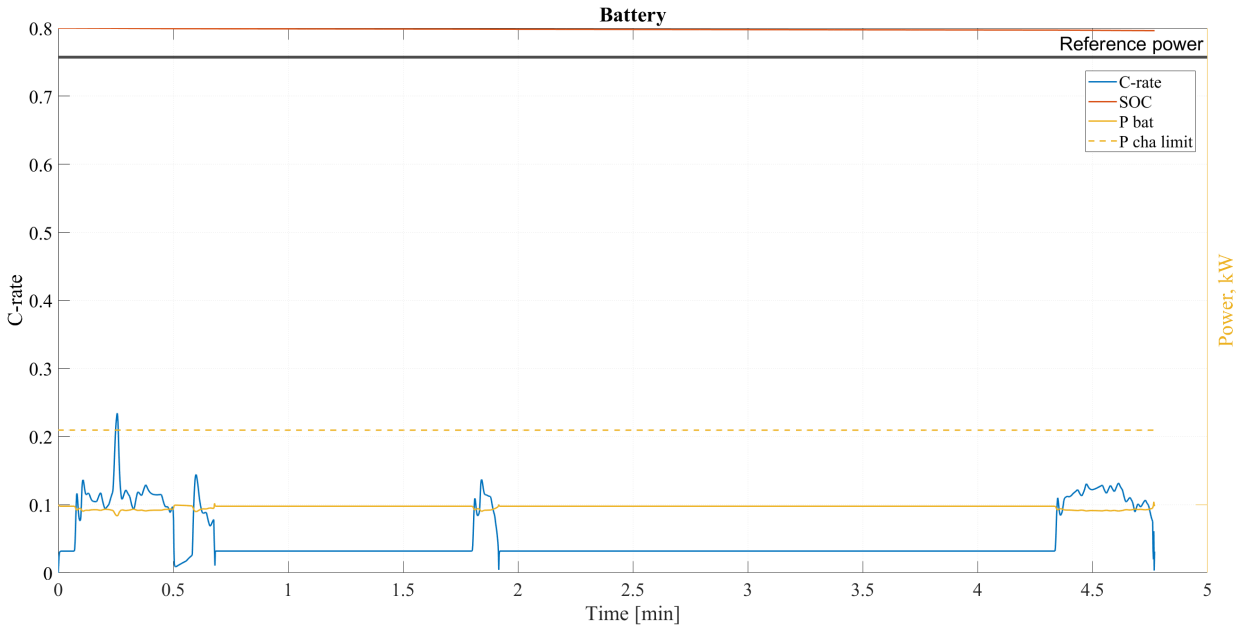


Figure 4.46: The battery behavior detected for the outlying drive cycle_196, BraveHawk_4, in the 33-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

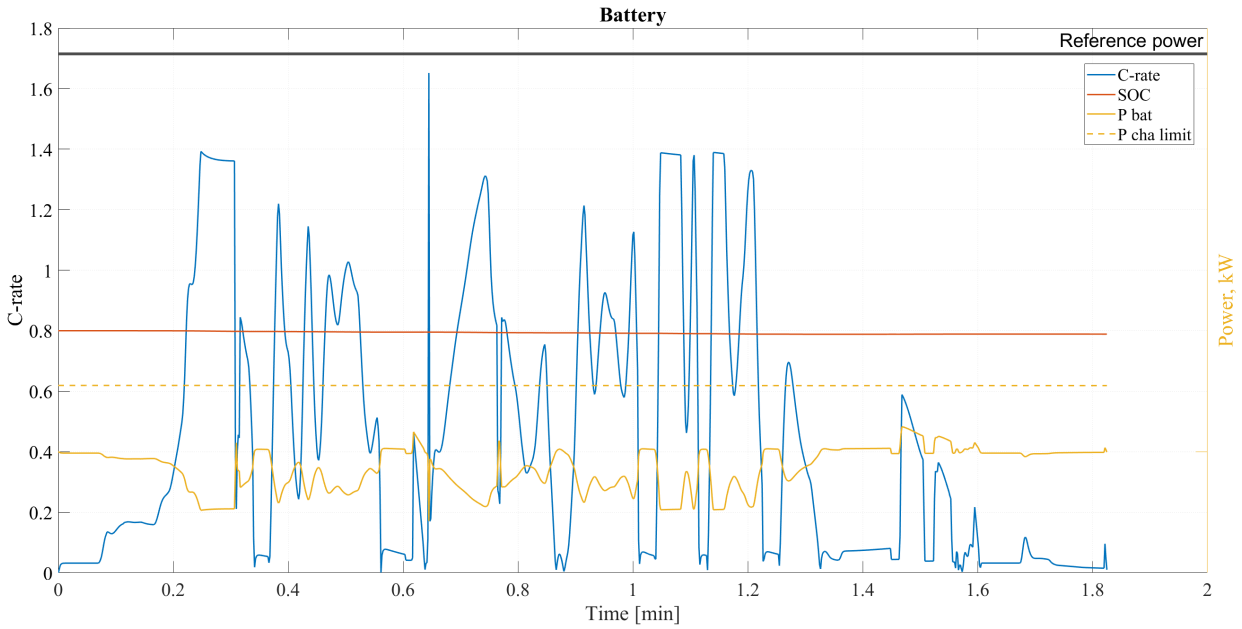


Figure 4.47: The battery behavior detected for the outlying drive cycle_906, ElegantPanther_6, in the 33-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

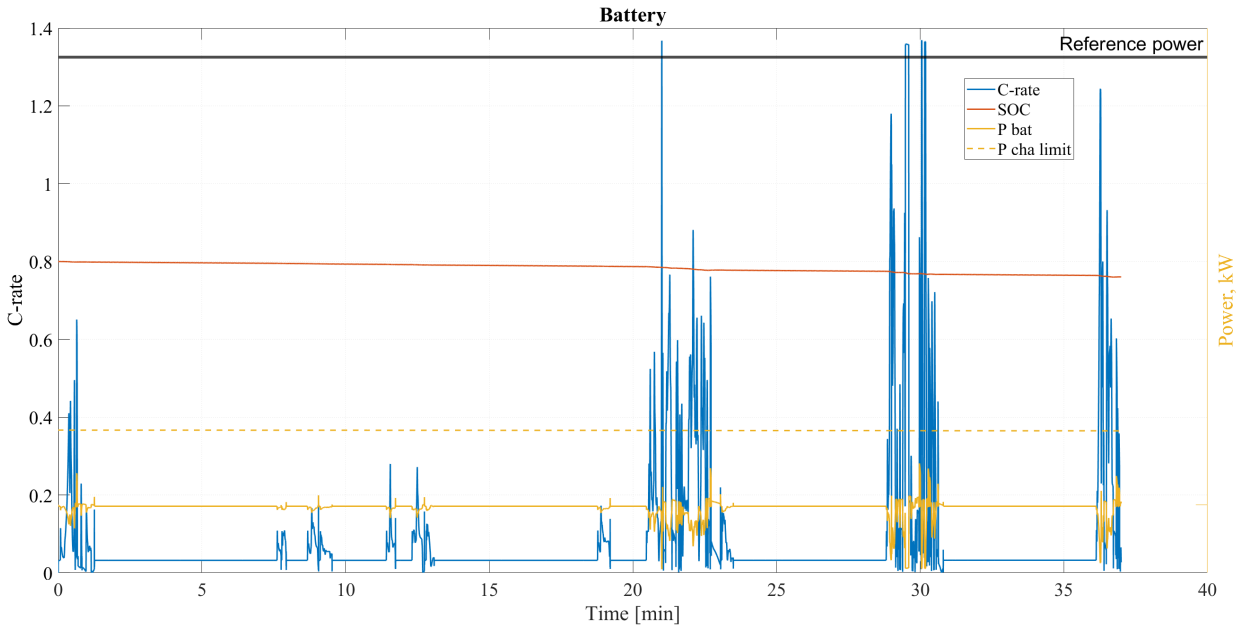


Figure 4.48: The battery behavior detected for the outlying drive cycle_1722, CleverWolf_7, in the 33-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

The SOC curves for the outlying drive cycles are primarily smooth and continuous for both vehicle configurations. Although there are pronounced short-term fluctuations their summed impact on the SOC levels remains limited. This indicates that transient battery loading for the outlying drive cycles, doesn't directly translate into unreasonable energy consumption, but rather variability within vehicle operation. The deviations are mainly a result from momentary changes in intensity and power demand, rather than different load operations.

Engine behavior

For the atypical drive cycles in Cluster 0, both vehicle configurations display a more inconsistent dispersion of the operating points when compared with the representative drive cycles and the operating points cover a wider range across the speed plane. For the 8-ton configuration the operating points remain concentrated at low torque levels, but more dispersed. In contrast the 33-ton configuration show a upward shift in operating points extending further towards the machine's maximum operating limits as a result from the increased mass. Figures 4.49-4.51 contains the 8-ton configuration and Figures 4.52-4.54 contains the 33-ton configuration.

Despite the increased scatter among the operating points the majority, for both configurations, remain within the moderate efficiency region of the efficiency map. This is also highlighted in Tables 4.30 and 4.31. The tables exhibit the mean-value efficiency of engine operation during the drive cycles. Short-lived instances towards lower-efficiency regions are present during abrupt peaks, mainly in the 33-ton configuration. Both traction and regenerative operating points remain within the boundaries of the torque and power limits for the engine and even for the most outlying drive cycles, there are no sustained operation near the max torque or regenerative boundaries.

Table 4.30: Electric machine efficiency data for the 8-ton simulations.

Parameter	Drive cycle_196	Drive cycle_906	Drive cycle_1722
EM 1, mean efficiency [%]	78.02	89.52	83.69
EM 2, mean efficiency [%]	76.57	88.61	84.22

Table 4.31: Electric machine efficiency data for the 33-ton simulations.

Parameter	Drive cycle_196	Drive cycle_906	Drive cycle_1722
EM 1, mean efficiency [%]	83.56	91.88	89.94
EM 2, mean efficiency [%]	80.40	91.29	90.02

Comparing the two vehicle configurations, the main distinction lies in the magnitude and duration of torque demand rather than the base vehicle operation. The 33-ton configuration exhibits larger magnitude of torque levels as well as more frequent operation near the higher efficiency regions, reflecting the increased load. The 8-ton configuration, while being more dispersed than the representative drive cycles, exhibits a denser operating distribution. For both cases, the main operation remains consistent, with the engine primarily utilized in high efficiency regions.

Although the drive cycles are distant in feature space, from the representative drive cycles, the engine behavior remains consistent with the defining characteristics of Cluster 0. The observed deviations are mainly a result from prolonged events and load variation rather than sustained high-speed or high-power operation. The observed dispersion in the operating points of the engine aligns with the clustering results presented earlier in Chapter 3.4, where Cluster 0 exhibited low intra-cluster variance. The engine efficiency maps provide a physical interpretation of this variance, emphasizing that even outlying drive cycles maintain efficient engine operation, with the exception of drive cycle_196 for BraveHawk_4.

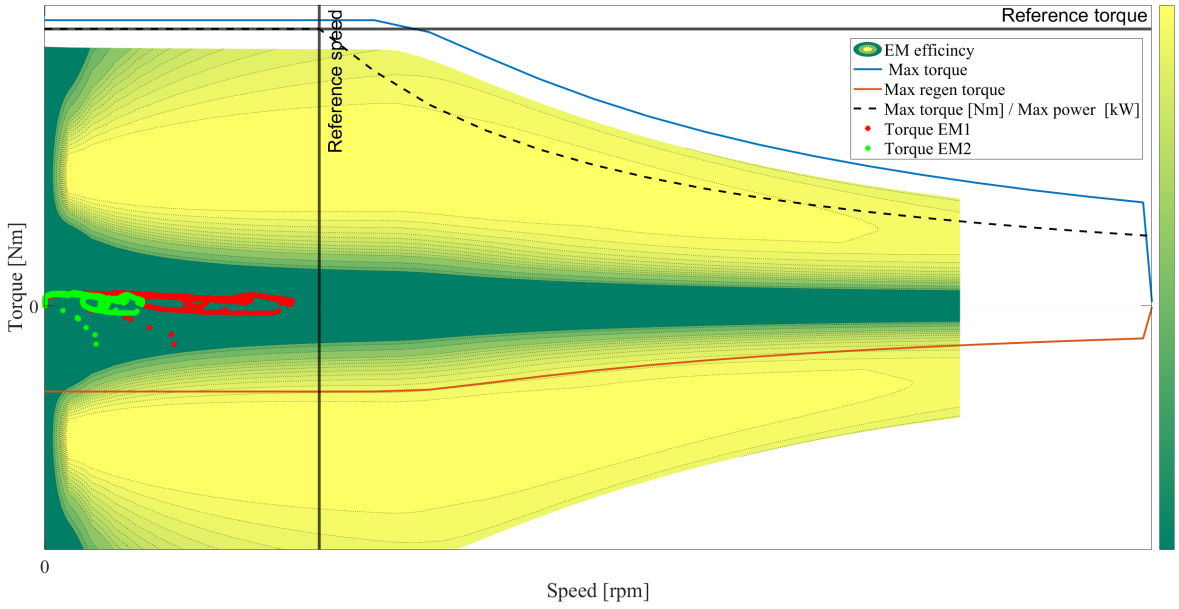


Figure 4.49: Engine behavior for the outlying drive cycle_196, BraveHawk_4, in the 8-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

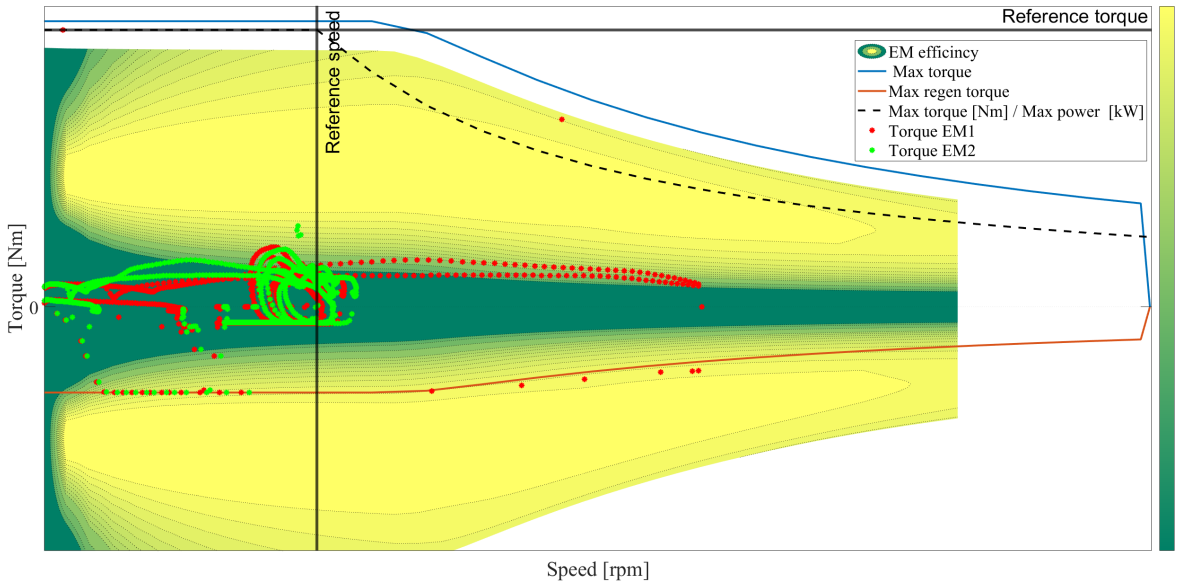


Figure 4.50: Engine behavior for the outlying drive cycle_906, ElegantPanther_6, in the 8-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

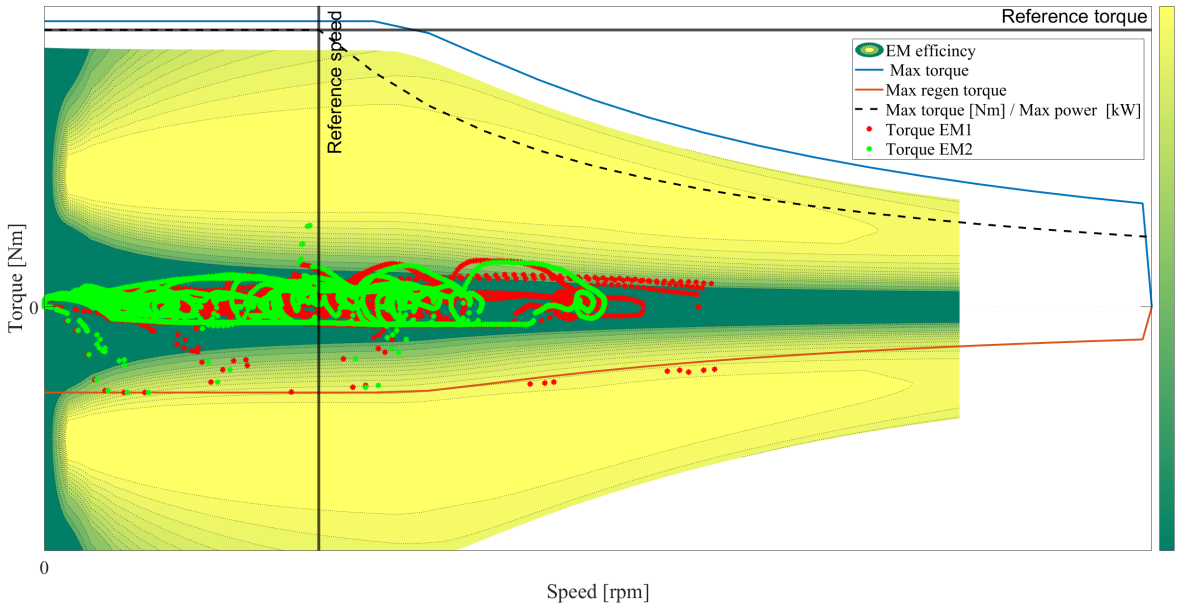


Figure 4.51: Engine behavior for the outlying drive cycle_1722, CleverWolf_7, in the 8-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

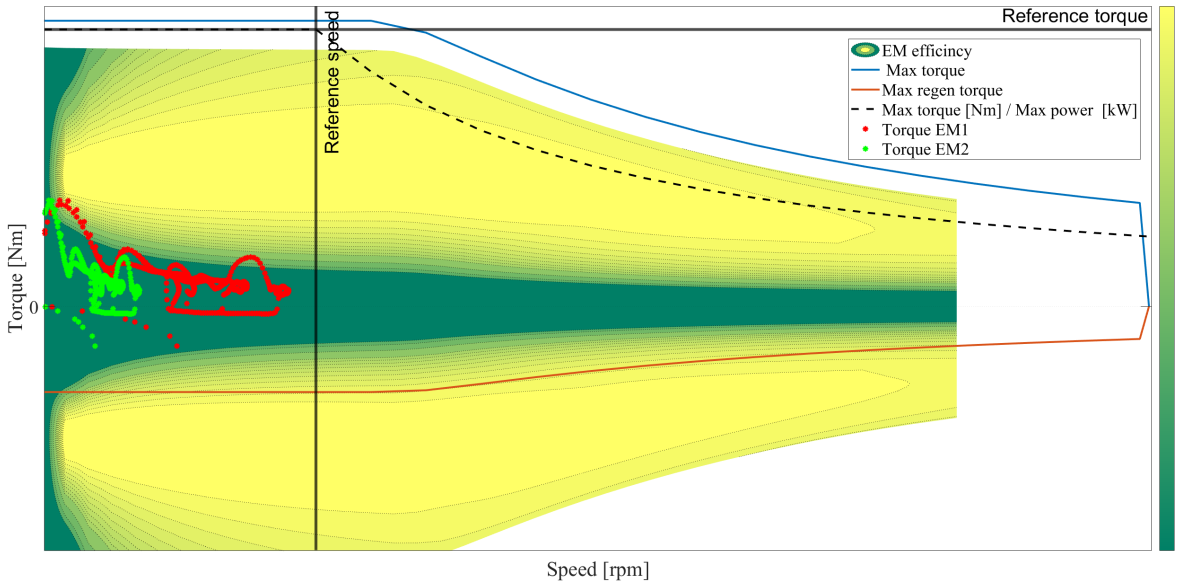


Figure 4.52: Engine behavior for the outlying drive cycle_196, BraveHawk_4, in the 33-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

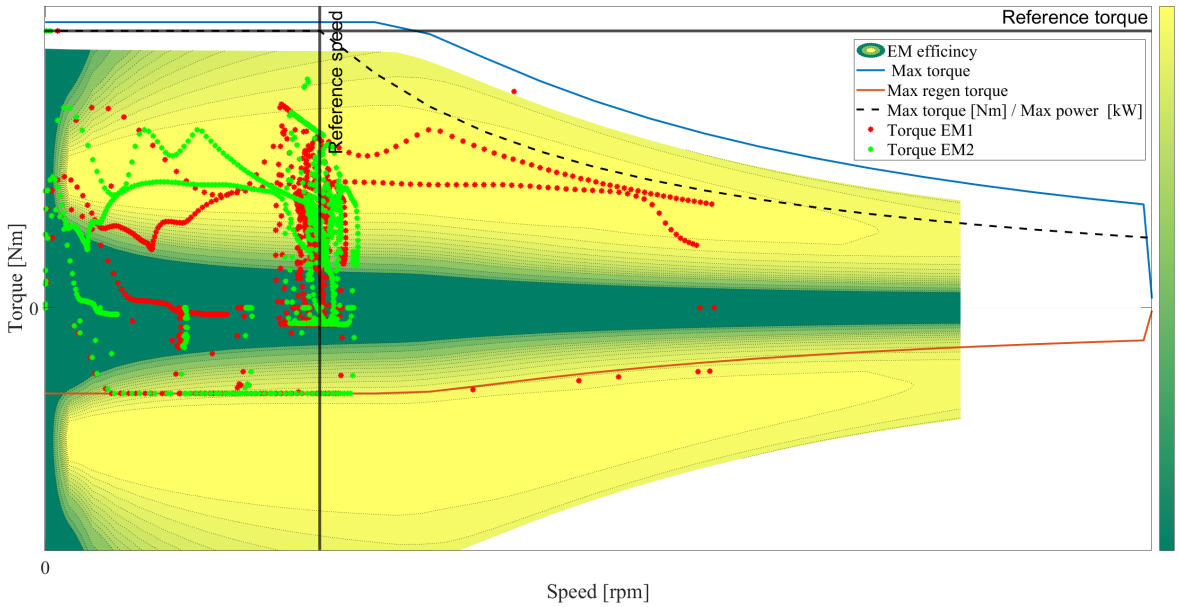


Figure 4.53: Engine behavior for the outlying drive cycle_906, ElegantPanther_6, in the 33-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

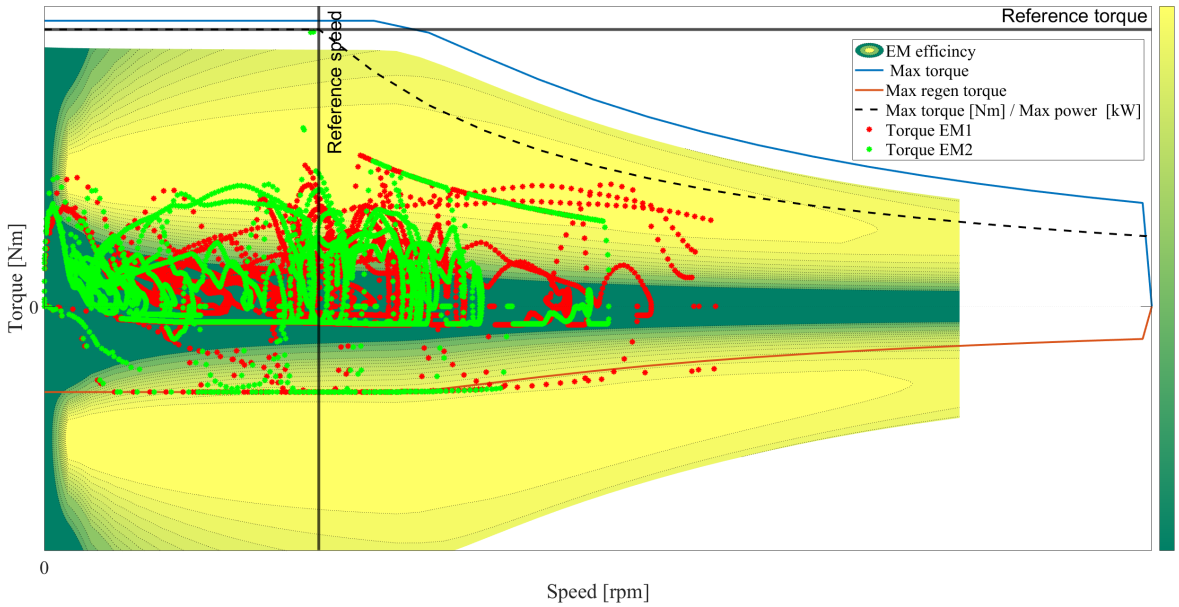


Figure 4.54: Engine behavior for the outlying drive cycle_1722, CleverWolf_7, in the 33-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

4.5 Drive cycles, Cluster 1

4.5.1 Representative drive cycles for Cluster 1

In order to describe the typical operational characteristics for Cluster 1, the representative drive cycles are analyzed. These cycles are the ones closest to the centroid of the cluster in feature space, and therefore reflect the most common operating conditions. The representative drive cycles are: drive cycle_135 for CleverWolf_7, drive cycle_271 for CleverWolf_7 and drive cycle_661 for CleverWolf_7. The Tables 4.32 and 4.33, highlight general information from both vehicle configurations for the representative drive cycles. One note to add is that *Travel time* and *Traveled distance* may vary between the two vehicle configurations as a result from approximations done during the calculations.

Table 4.32: General information for the 8-ton simulations.

Parameter	Drive cycle_135	Drive cycle_271	Drive cycle_661
Travel time [min]	47	45	52
Traveled distance [km]	4.0401	7.1712	7.0531
Average speed [km/h]	5.1861	9.5993	8.1111
Energy cons [kWh/100km]	118.6	84.4	88.8
Electric cons [kWh]	5.1	6.5	6.7

Table 4.33: General information for the 33-ton simulations.

Parameter	Drive cycle_135	Drive cycle_271	Drive cycle_661
Travel time [min]	46	43	50
Traveled distance [km]	4.0401	7.1723	7.0538
Average speed [km/h]	5.2941	9.9255	8.4428
Energy cons [kWh/100km]	174.6	172.5	163.6
Electric cons [kWh]	7.6	13.2	12.3

When comparing the representative drive cycles in Cluster 1 to the representative drive cycles in Cluster 0, Cluster 1 exhibit slightly longer drive cycles. These drive cycles cover a larger travel distance, a higher average speed, and a larger total energy consumption. The distance normalized energy consumption for the representative Cluster 1 drive cycles is larger, indicating heavier drive cycle operations.

Electromagnetic torque and speed

Under the operating condition for the 8-ton configuration frequent and repetitive traction events, can be seen Figures 4.55-4.57. The traction events are characterized by changes in torque, between low to moderate levels during propulsion and moderate negative torque values during braking and regenerative operation. Although Figures 4.58-4.60 exhibit a different vehicle configuration, the torque and speed profiles are similar, with frequent operating events followed by an idle period rather than extended continuous operation. The corresponding speed values also indicates an operational pattern consisting of repeated acceleration phases, increasing the rotational speed of the engine, followed by short near-zero periods. Both vehicle configurations, see Figures 4.55-4.57 and 4.58-4.60, highlight engine one reaching higher peak-values for speed, once again due to more gear ratios as previously mentioned in subsection 4.4.1.

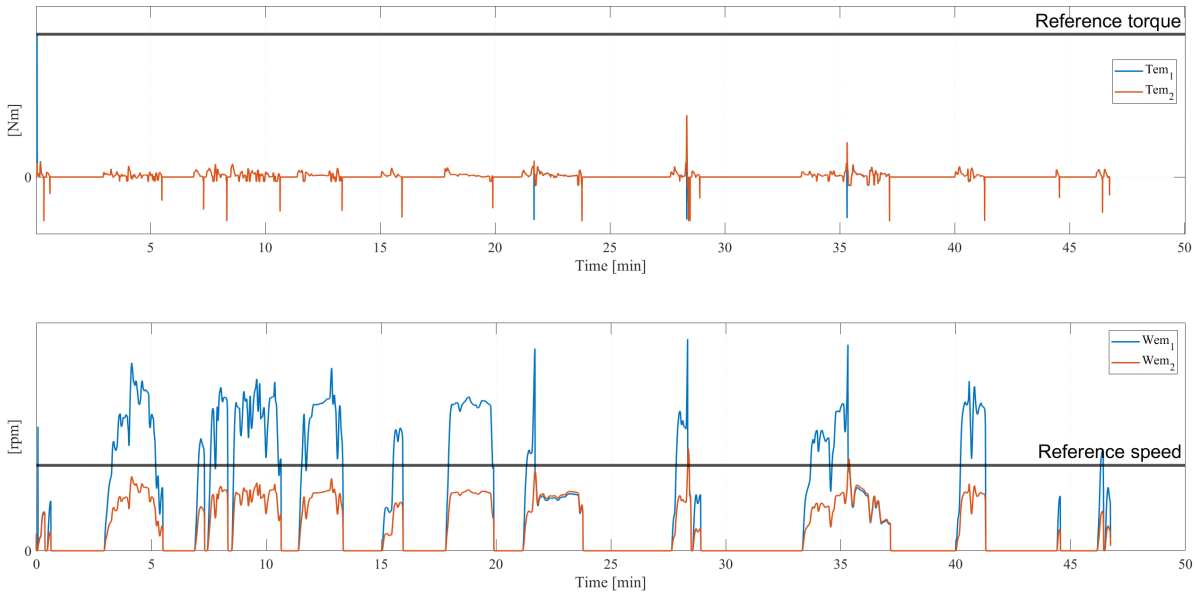


Figure 4.55: Electromagnetic torque and speed for representative drive cycle_135, CleverWolf_7, in the 8-ton configuration for Cluster 1.

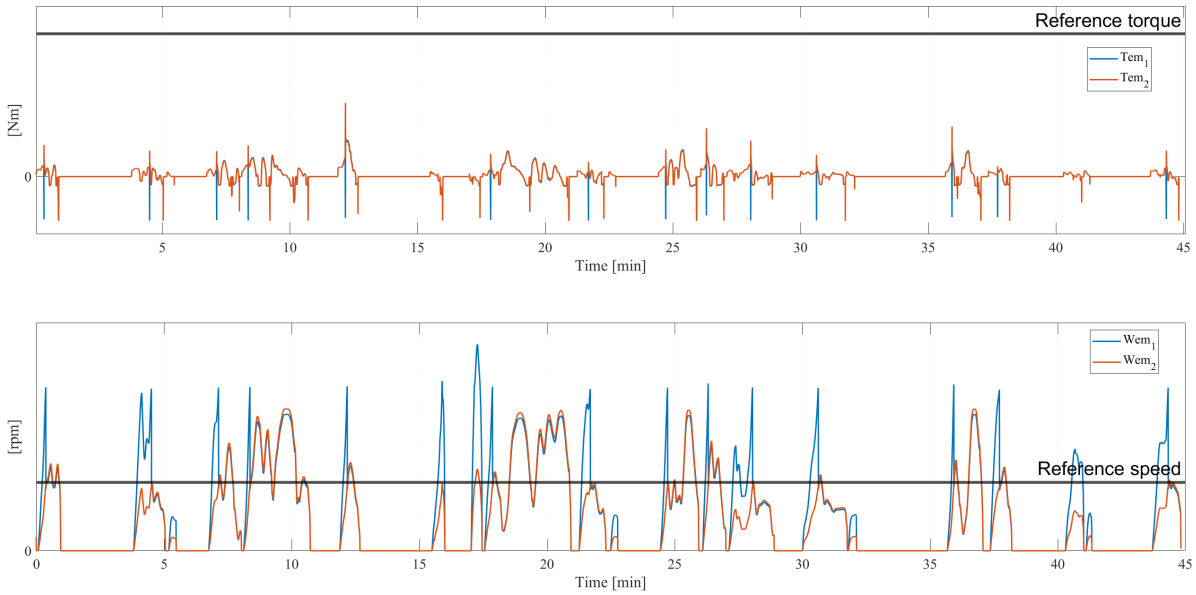


Figure 4.56: Electromagnetic torque and speed for representative drive cycle_271, CleverWolf_7, in the 8-ton configuration for Cluster 1.

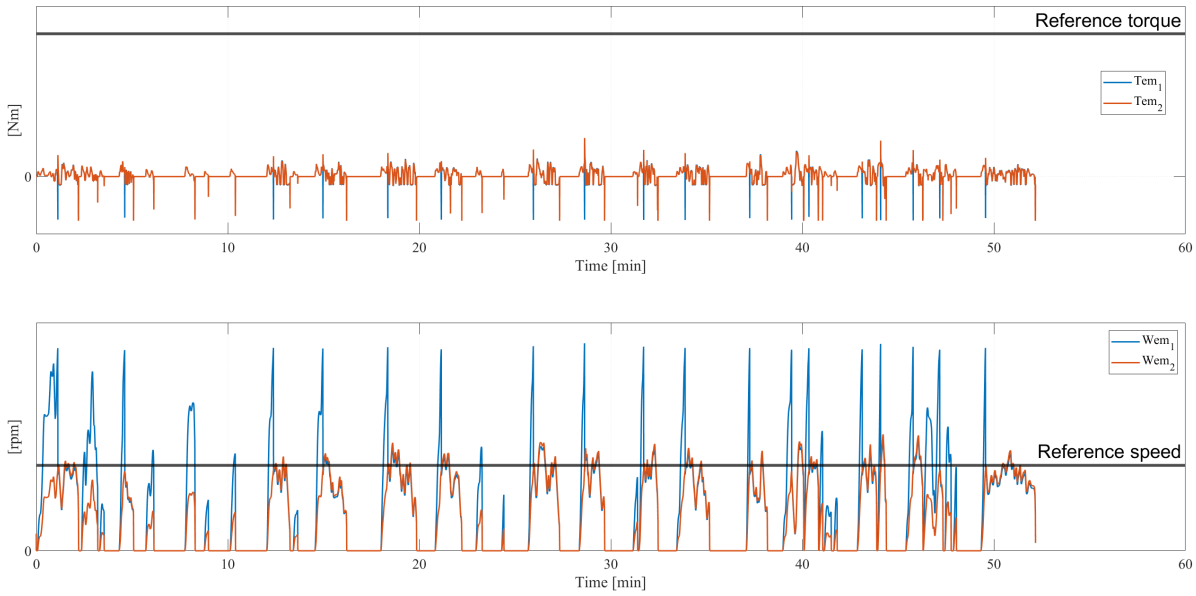


Figure 4.57: Electromagnetic torque and speed for representative drive cycle_661, CleverWolf_7, in the 8-ton configuration for Cluster 1.

For the 33-ton configuration the drive cycles, in Figures 4.58-4.60, clearly exhibit an increased torque activity. When comparing drive cycle_661 with both drive cycle_271 and drive cycle_135 for CleverWolf_7, larger torque magnitude levels are present as well as clearer fluctuations between the traction events. Both engines contribute actively with a noticeable torque variation reflecting the increased demand resulting from the increased vehicle mass, when compared to the 8-ton configuration. Negative torque events are more pronounced as well, indicating the larger demand in braking- and regenerative power associated with increased vehicle mass and inertia. Lastly comparing Figures 4.55-4.57 and 4.58-4.60, the operating behavior exhibits balanced engine operation regardless of vehicle mass.

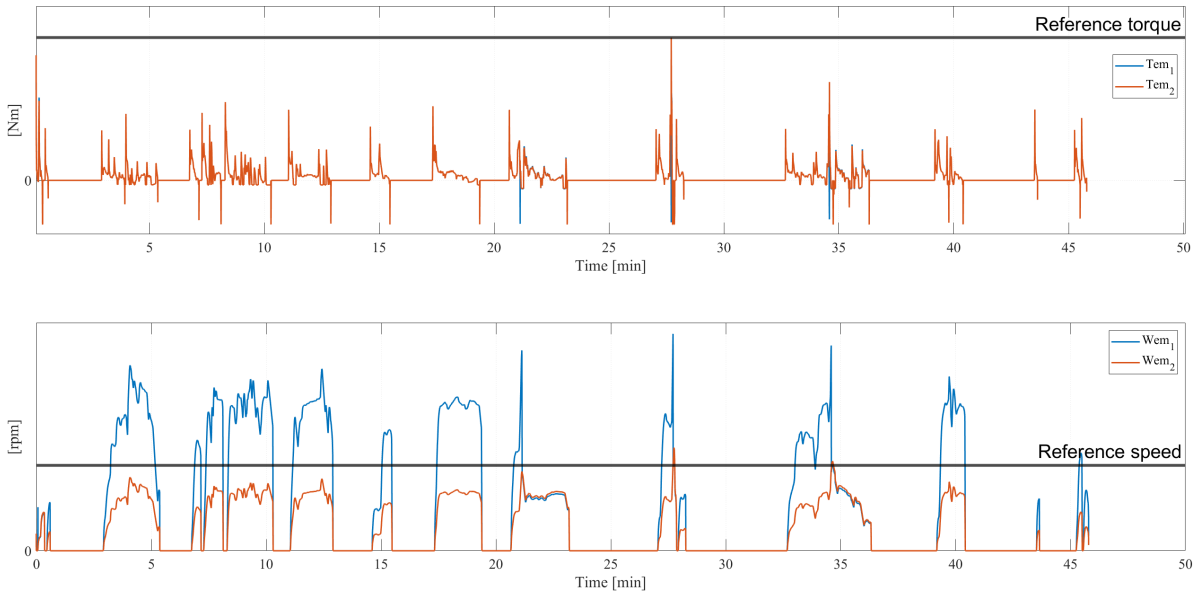


Figure 4.58: Electromagnetic torque and speed for representative drive cycle_135, CleverWolf_7, in the 33-ton configuration for Cluster 1.

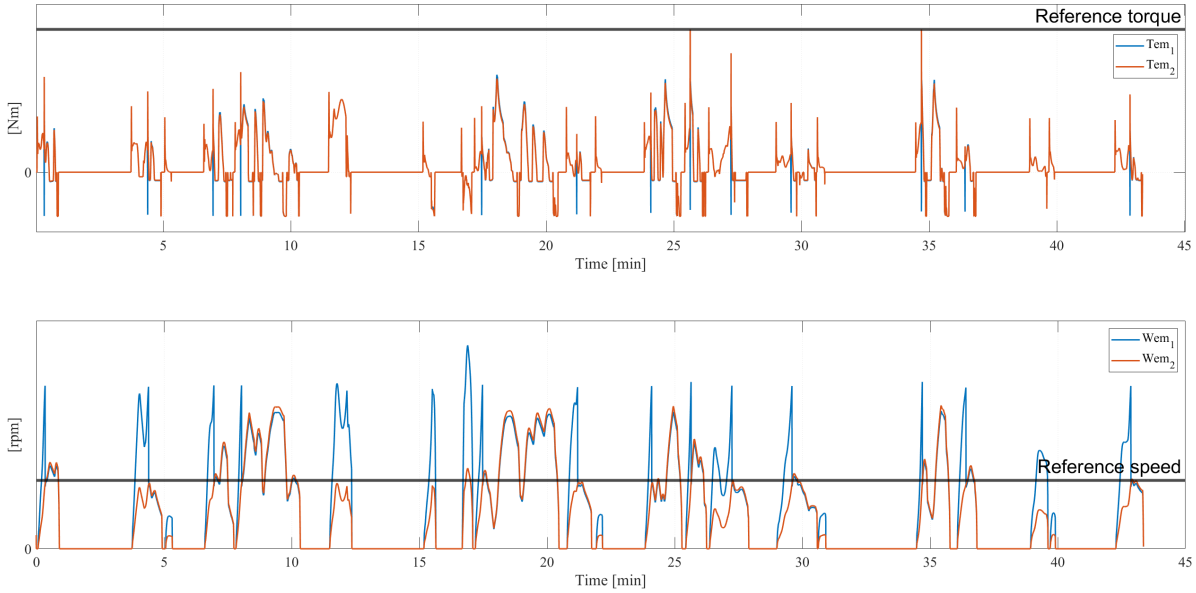


Figure 4.59: Electromagnetic torque and speed for representative drive cycle_271, CleverWolf_7, in the 33-ton configuration for Cluster 1.

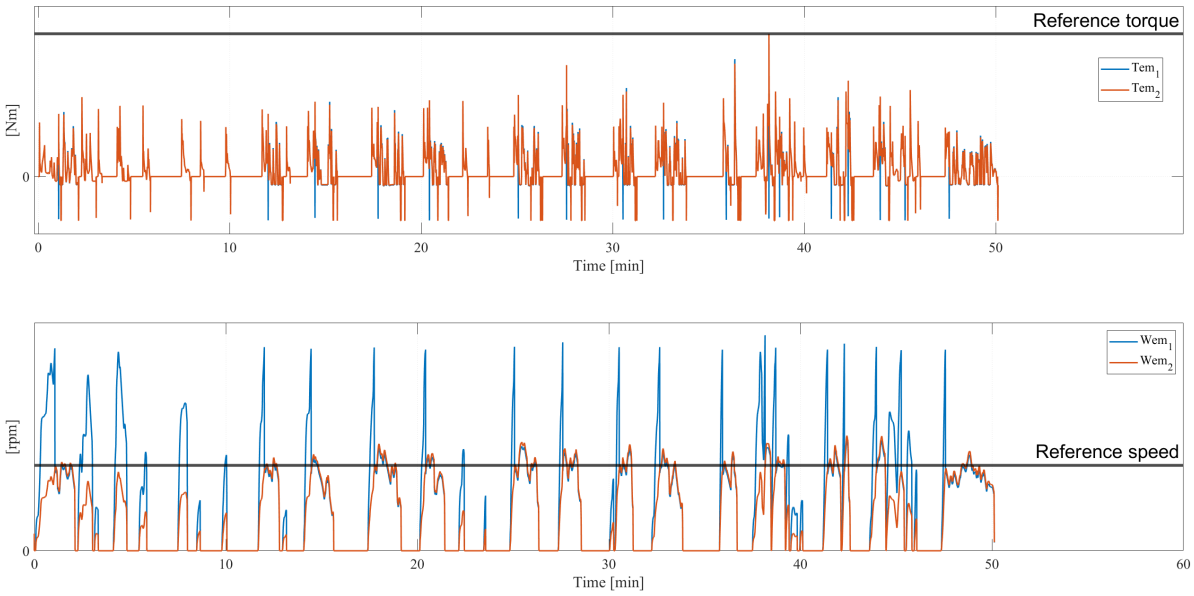


Figure 4.60: Electromagnetic torque and speed for representative drive cycle_661, CleverWolf_7, in the 33-ton configuration for Cluster 1.

Battery behavior

For the representative drive cycles in Cluster 1, regardless of mass configuration, the power demand from the battery fluctuates. This is exhibited as frequent short-lived C_{rate} peaks in Figures 4.61-4.63 for the 8-ton configuration and Figures 4.64-4.66 for the 33-ton configuration. These high peaks are consistent with the start-stop behavior observed in the operational driving pattern.

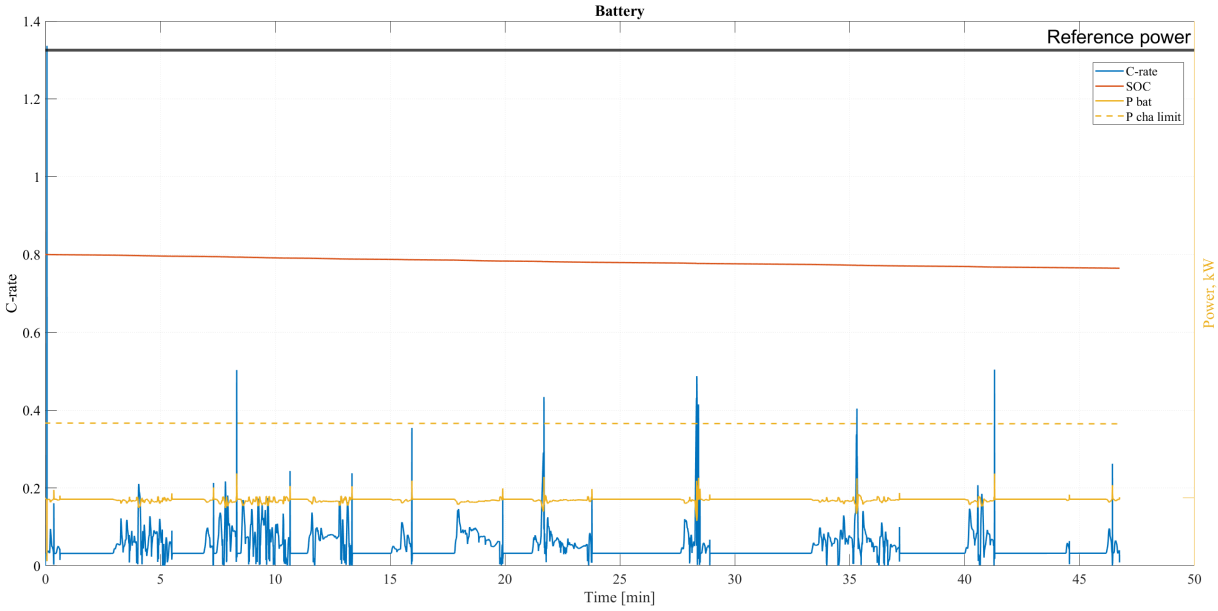


Figure 4.61: The battery behavior detected for the representative drive cycle_135, CleverWolf_7, in the 8-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

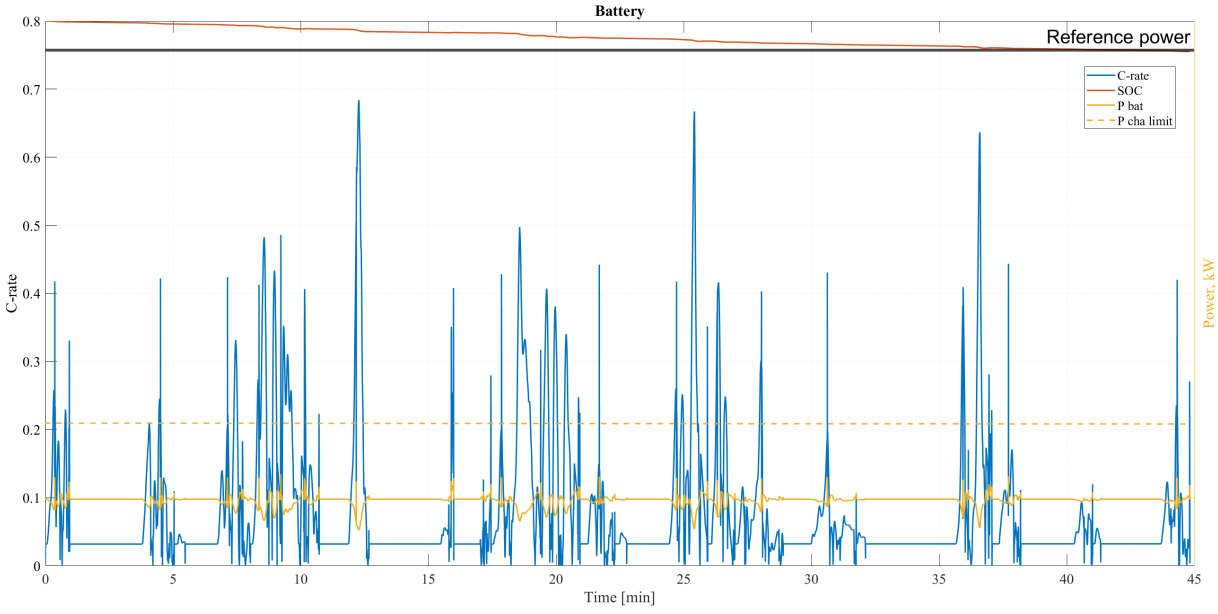


Figure 4.62: The battery behavior detected for the representative drive cycle_271, CleverWolf_7, in the 8-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

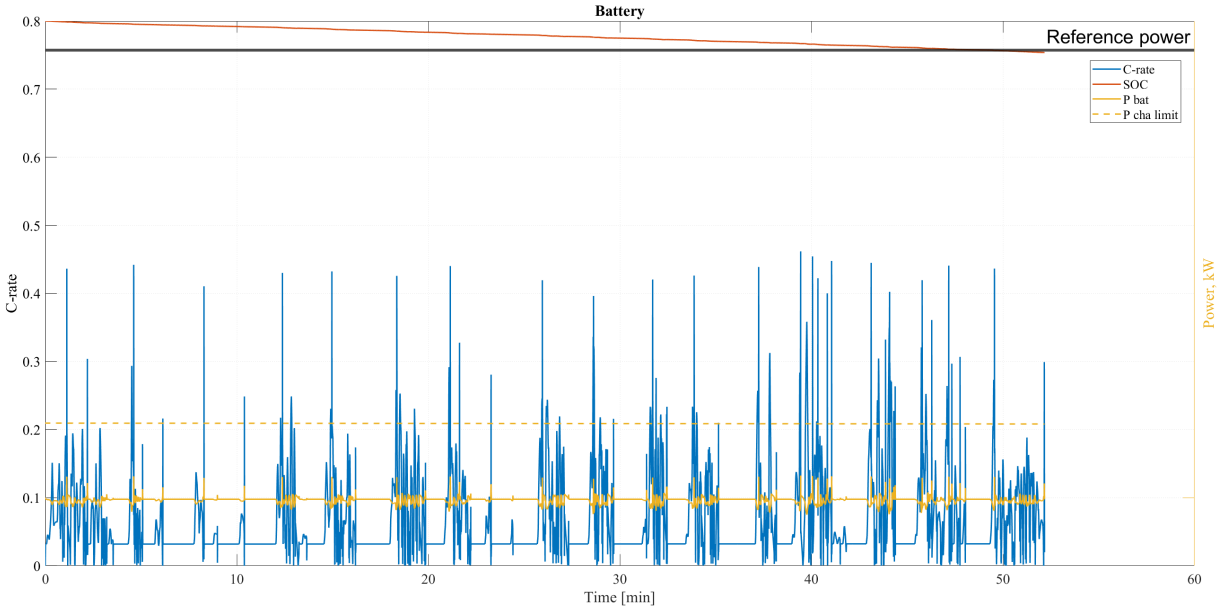


Figure 4.63: The battery behavior detected for the representative drive cycle_661, CleverWolf_7, in the 8-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

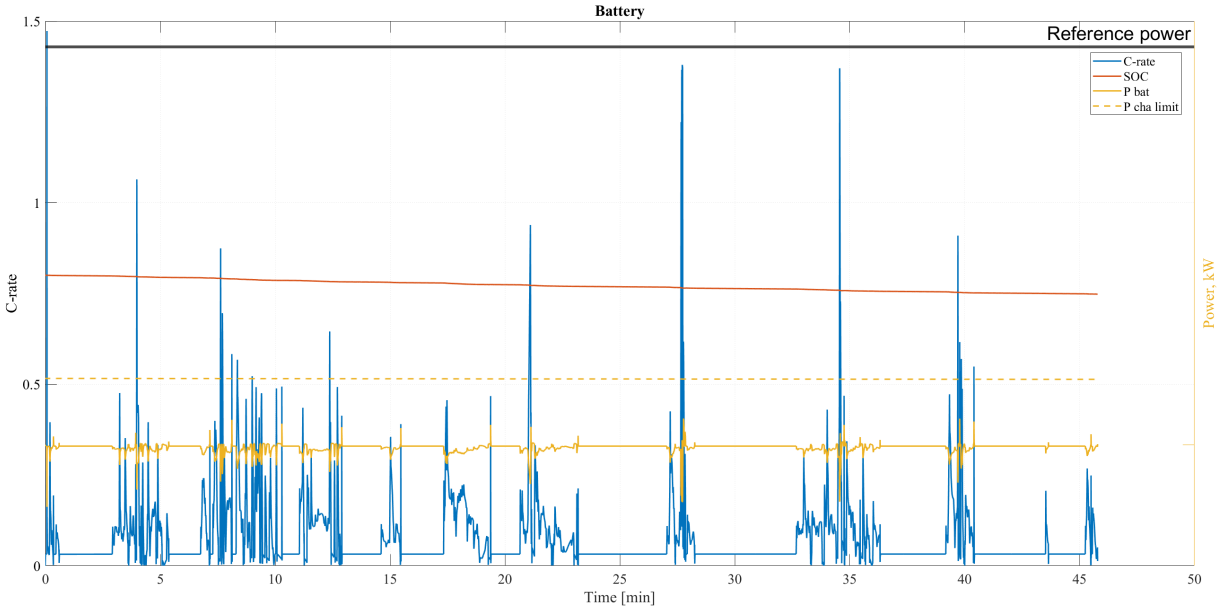


Figure 4.64: The battery behavior detected for the representative drive cycle_135, CleverWolf_7, in the 33-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

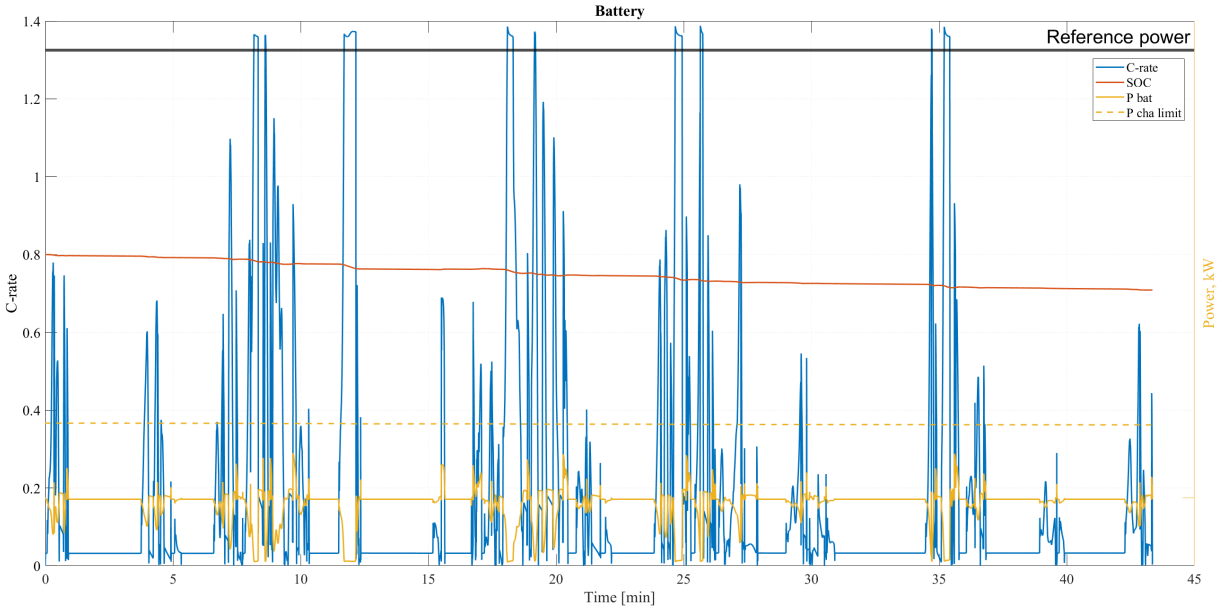


Figure 4.65: The battery behavior detected for the representative drive cycle_271, CleverWolf_7, in the 33-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

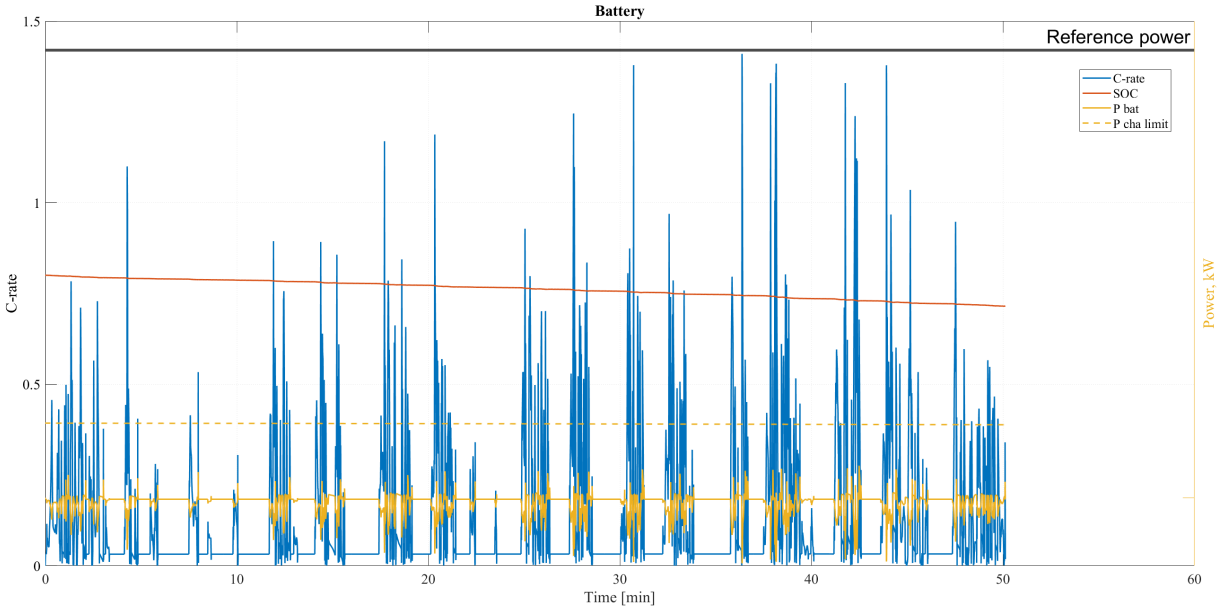


Figure 4.66: The battery behavior detected for the representative drive cycle_661, CleverWolf_7, in the 33-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

For the configuration of 8-ton, almost every peak occur at low to moderate C_{rates}. For the 33-ton configuration the C_{rate} peaks occur at much larger amplitude, reflecting how the increased mass directly affects the battery power demand, rather than the usage pattern between the drive cycles. This behavior indicates that the battery stress is dominated by short-lived transient events of small power magnitude. However, although both configurations display battery power peaks they are short-lived and doesn't translate into sustained operation near the charging- or power limit.

The SOC path progresses smoothly for both configurations, during the drive cycle with out abrupt depletion. Even though there is a clear difference in mass, the 33-ton configuration only exhibits a slightly steeper SOC decrease for drive cycle_661. The smooth SOC path indicates that despite frequent power transients there is near to none/limited energy impact during the operation. The rapid power exchange, as indicated by the short-lived peaks in the C_{rate}-graph, results in high variability but low cumulative stress on the battery during the drive cycle.

The observed battery behavior for Cluster 1 agrees with the previously clustered results stating that Cluster 1 consists of high torque variation, rather than sustained power- and/or energy demand. These results are consistent with the presented battery dynamics. Short-lived power pulses have limited influence on the SOC when compared to long-lived high-load operation. Accordingly, different vehicle mass mainly influences the power demand rather than the battery usage during operation.

The Tables 4.34 and 4.35 presents higher total captured- and output energy than for the representative drive cycles in Cluster 0. The regenerated energy is more pronounced for Cluster 1, with the exception of drive cycle_135 for the 8-ton configuration. Cluster 1 consistently reaches higher energy throughput values while maintaining an ESS-efficiency above 98%.

Table 4.34: Energy storage data for the 8-ton simulations.

Parameter	**Drive cycle_135**	**Drive cycle_271**	**Drive cycle_661**
Integrated current [Ah]	7.24	9.15	9.45
Total Energy Output [kWh]	5.3	7.3	7.3
Total Energy captured [kWh]	0.12	0.85	0.65
ESS efficiency [%]	99.8	99.6	99.7
Energy throughput [kWh/h]	6.91	10.91	9.19
End SOC [%]	76.47	75.54	75.39
SOC difference [%-points]	3.53	4.46	4.61
Energy charge station [kWh]	0	0	0

Table 4.35: Energy storage data for the 33-ton simulations.

Parameter	Drive cycle_ 135	Drive cycle_271	Drive cycle_661
Integrated current [Ah]	10.66	18.69	17.43
Total Energy Output [kWh]	7.8	15.8	13.7
Total Energy captured [kWh]	0.27	2.89	1.51
ESS efficiency [%]	99.5	98.5	99.2
Energy throughput [kWh/h]	10.57	25.91	18.21
End SOC [%]	74.81	70.89	71.50
SOC difference [%-points]	5.19	9.11	8.50
Energy charge station [kWh]	0	0	0

Engine behavior

Figures 4.67-4.69 shows the engine behavior for the 8-ton configuration and Figures 4.70-4.72 for the 33-ton configuration. The Figures highlight a vastly spread in operating points for the engines in the torque-speed plane. This reflects the frequent changes in operation during the drive cycles. Both vehicle configurations display denser operating points at low and moderate torque levels across a broad speed range. Both vehicle configurations operate above 70% efficiency. The observed distribution between positive and negative torque indicates changes between traction demand and regenerative braking events. The events are mainly occurring at low and moderate speeds, where the traction demand is consistent with start-stop driving conditions. The Tables 4.36 and 4.37, highlight the mean-efficiency value of operation for both vehicle configurations during the drive cycles.

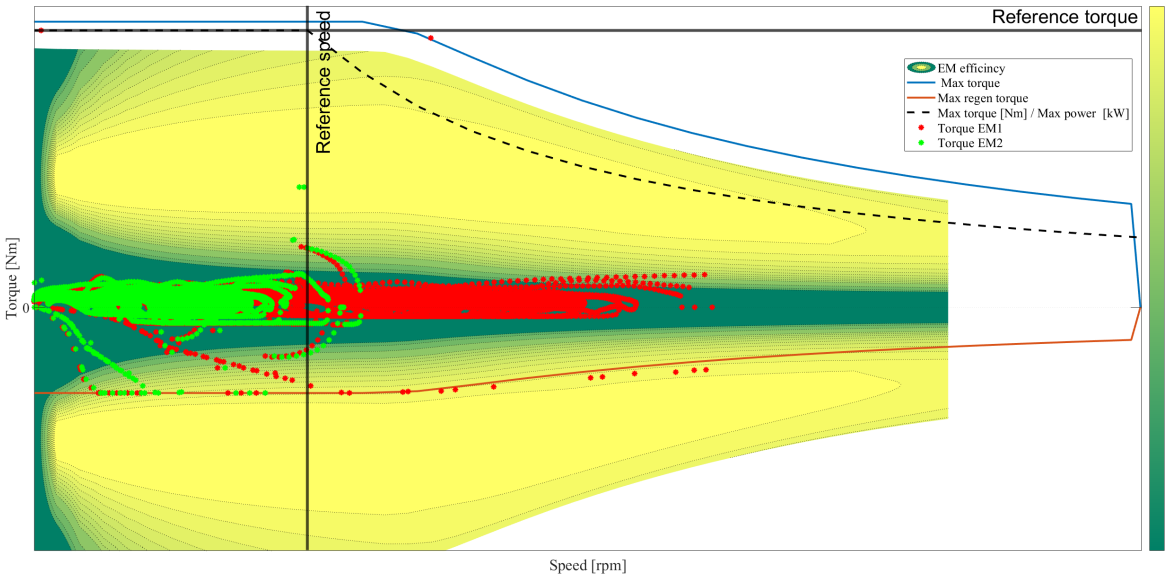


Figure 4.67: Engine behavior for the representative drive cycle_135, CleverWolf_7, in the 8-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

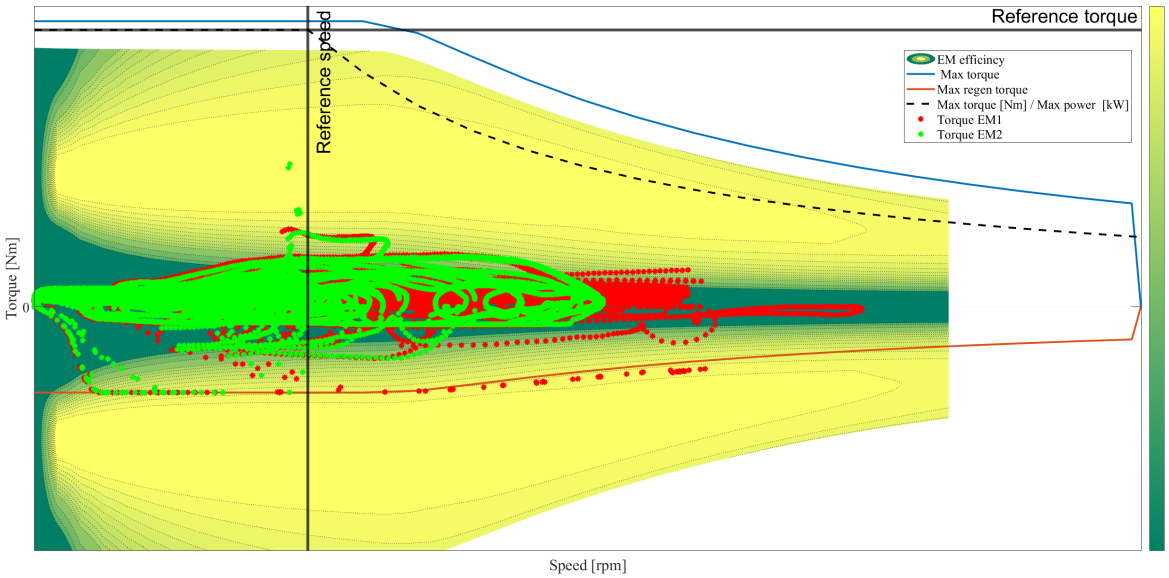


Figure 4.68: Engine behavior for the representative drive cycle_271, CleverWolf_7, in the 8-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

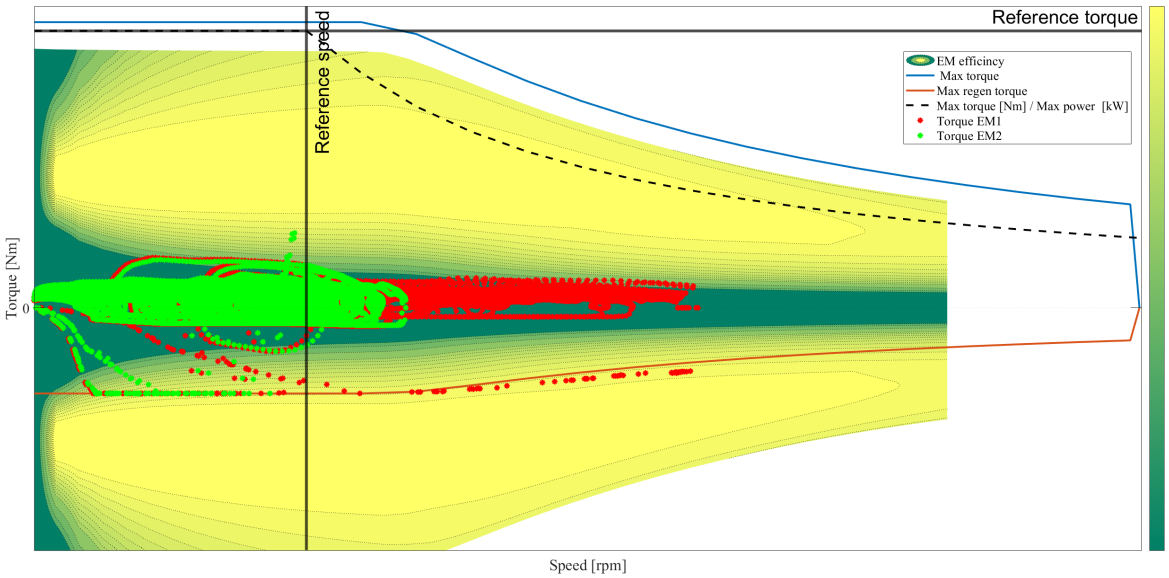


Figure 4.69: Engine behavior for the representative drive cycle_661, CleverWolf_7, in the 8-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

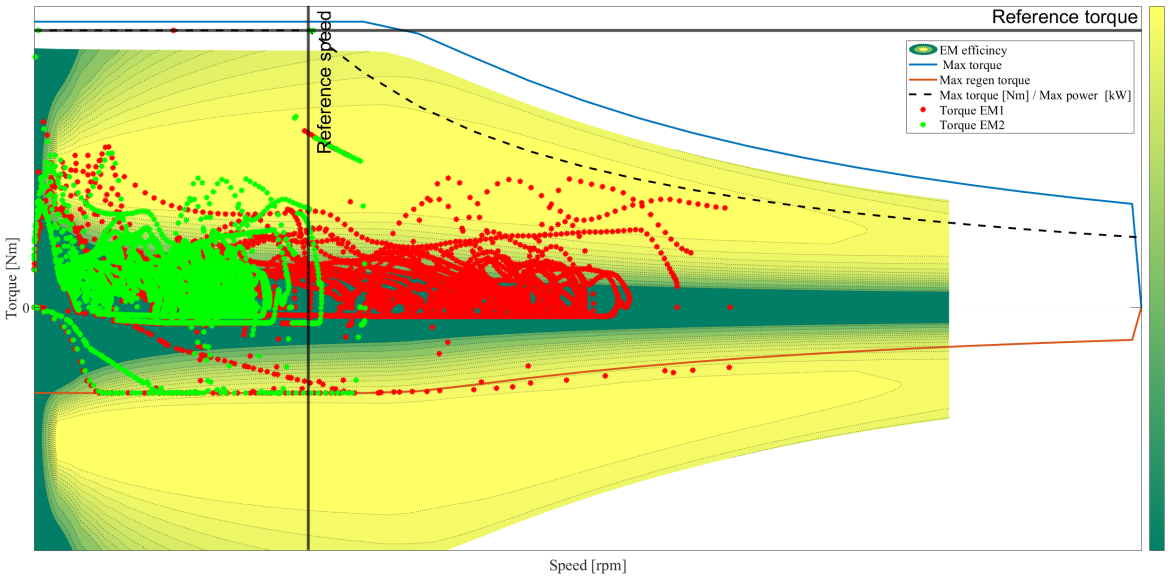


Figure 4.70: Engine behavior for the representative drive cycle_135, CleverWolf_7, in the 33-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

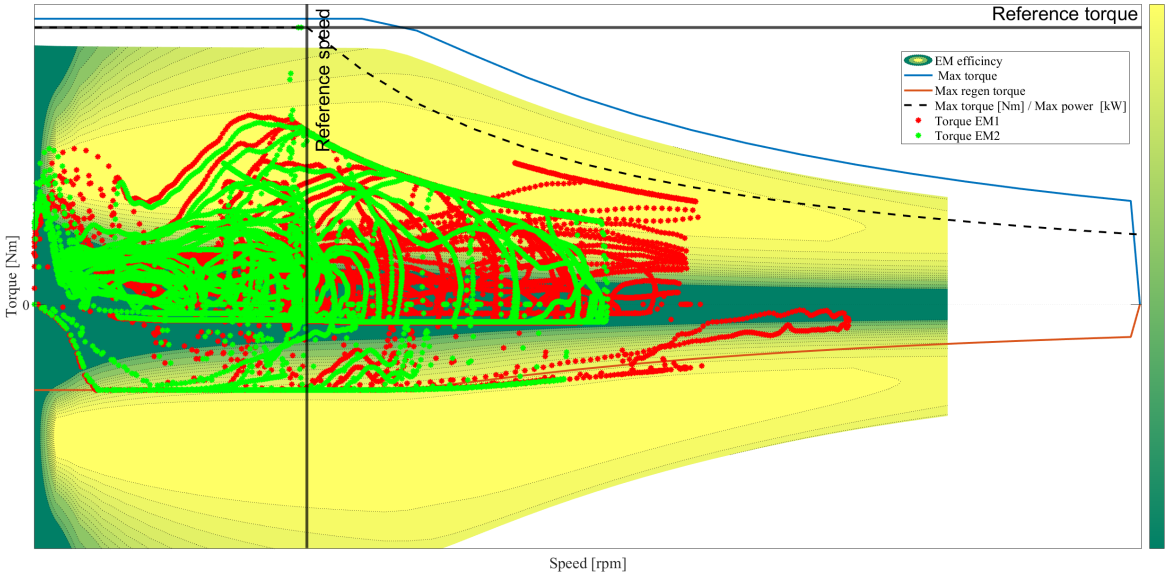


Figure 4.71: Engine behavior for the representative drive cycle_271, CleverWolf_7, in the 33-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

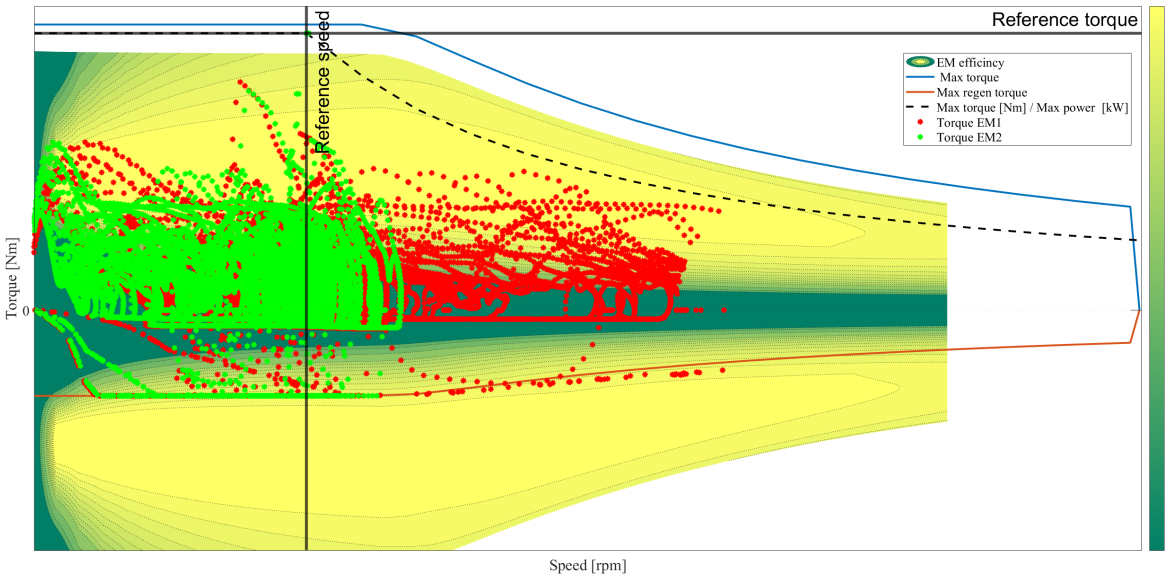


Figure 4.72: Engine behavior for the representative drive cycle_661, CleverWolf_7, in the 33-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

Table 4.36: Electric machine efficiency data for the 8-ton simulations.

Parameter	Drive cycle_135	Drive cycle_271	Drive cycle_661
EM 1, mean efficiency [%]	75.47	85.03	85.24
EM 2, mean efficiency [%]	80.09	85.32	85.20

Table 4.37: Electric machine efficiency data for the 33-ton simulations.

Parameter	Drive cycle_135	Drive cycle_271	Drive cycle_661
EM 1, mean efficiency [%]	82.89	91.40	89.87
EM 2, mean efficiency [%]	85.72	91.15	89.72

When comparing the 8-ton configuration with the 33-ton configuration the increased mass mainly results in an increased torque demand. Following the increased torque demand there is a noticeable shift in operating points towards larger levels of torque amplitude. However, despite the shift, the distribution of operating points remains similar between the two vehicle configurations. Both configurations lie within their torque limits. Some operating points are present at the torque limit but are short-lived and therefore doesn't contribute to sustained high-load operation.

Across both configurations the operating points are at their densest at moderate to high engine efficiency. There are a few instances of low-efficiency operating points, at low torque levels. These operating points are most likely due to transient changes in torque and speed and have limited influence on the overall efficiency during the drive cycle.

Both configurations clearly exhibit negative torque operating points, confirming the existence of regenerative braking. The regenerative braking events are spread across a wide range of speed levels. There are instances of operating points where the regenerative torque levels reach the set limit and some operating points are even beyond it. However, due to their transient nature and few occurrences, their impact on the engine operation will be limited. The similar regenerative braking shapes for the two vehicle configurations indicate that the main difference is due to the increased mass and not due to different driving behavior.

When combining the presented results, the engine operating points indicate that Cluster 1 represents drive cycles with higher torque demand and frequent torque changes. The engine is mainly operating at low and moderate torque levels, resulting in a wide spread of operating points across the speed plane. A few, but present, extreme operating points don't pose a direct danger for thermal or mechanical malfunction. The observed torque-speed operating points are consistent with previous results, see Chapter 3.4, where Cluster 1 was characterized by high torque variance. The energy losses for the drive cycles are mainly driven by abrupt torque reversals and low speed operation. The shift in torque operating points, between the 8-ton- and 33-ton configuration, is a direct consequence of the increased mass rather than a change in drive cycle operation for Cluster 1.

4.5.2 Outlying drive cycles for Cluster 1

In order to describe atypical operational characteristics for Cluster 1, the outlying drive cycles are analyzed. These drive cycles are the ones farthest away from the centroid of the cluster in feature space, and therefore reflect the most extreme operating conditions within the cluster. The outlying drive cycles are: drive cycle_217 for BraveHawk_4, drive cycle_331 for ElegantPanther_6 and drive cycle_1207 for CleverWolf_7.

One thing to add is that drive cycle_217, for BraveHawk_4, has too few non-zero values in order to generate a road cycle. Therefore, it is classified as idle. There is no real interest in examining this drive cycle since it doesn't generate any data, hence it won't be mentioned further in this section. With this in mind, the Tables 4.38 and 4.39, present general information for drive cycle_331 and drive cycle_1207 during operation. One note to add is that *Travel time* and *Traveled distance* may vary between the two vehicle configurations as a result from approximations done during the calculations. From the tables it can be noticed that the outlying drive cycles display very short drive cycles. Although short-lived, the drive cycles cover a lot of distance. The two vehicle configurations display a similar amount of distance-normalized energy consumption, both lower than the representative drive cycles in Cluster 1.

Table 4.38: General information for the 8-ton simulations.

Parameter	**Drive cycle_331**	**Drive cycle_1207**
Travel time [min]	14	13
Traveled distance [km]	2.4517	4.4697
Average speed [km/h]	10.6109	20.4981
Energy cons [kWh/100km]	76.3	66.1
Electric cons [kWh]	2.0	3.2

Table 4.39: General information for the 33-ton simulations.

Parameter	Drive cycle_331	Drive cycle_1207
Travel time [min]	13	13
Traveled distance [km]	2.4517	4.4694
Average speed [km/h]	11.1355	21.1905
Energy cons [kWh/100km]	151.5	176.9
Electric cons [kWh]	4.0	8.5

Electromagnetic torque and speed

Figures 4.73-4.74 exhibit smoother propulsion phases when compared to Figures 4.67-4.69, clearest seen when looking at drive cycle_1207 for CleverWolf_7. During acceleration both drive cycles exhibit a more even and coherent speed profile also seen in Figures 4.75-4.76. The speed profiles are also more aligned with fewer, but still present, peak speed values for engine one when compared to Figures 4.67-4.69.

The variation in torque is moderate, reflecting stable operation for both outlying drive cycles in Cluster 1, despite being atypical. Although not as start-stop dominated, as the representative drive cycles for Cluster 1, Figures 4.73- 4.74 exhibits clear idle periods separating propulsion events. Negative torque values are rare, indicating less braking and regeneration power events during the drive cycles.

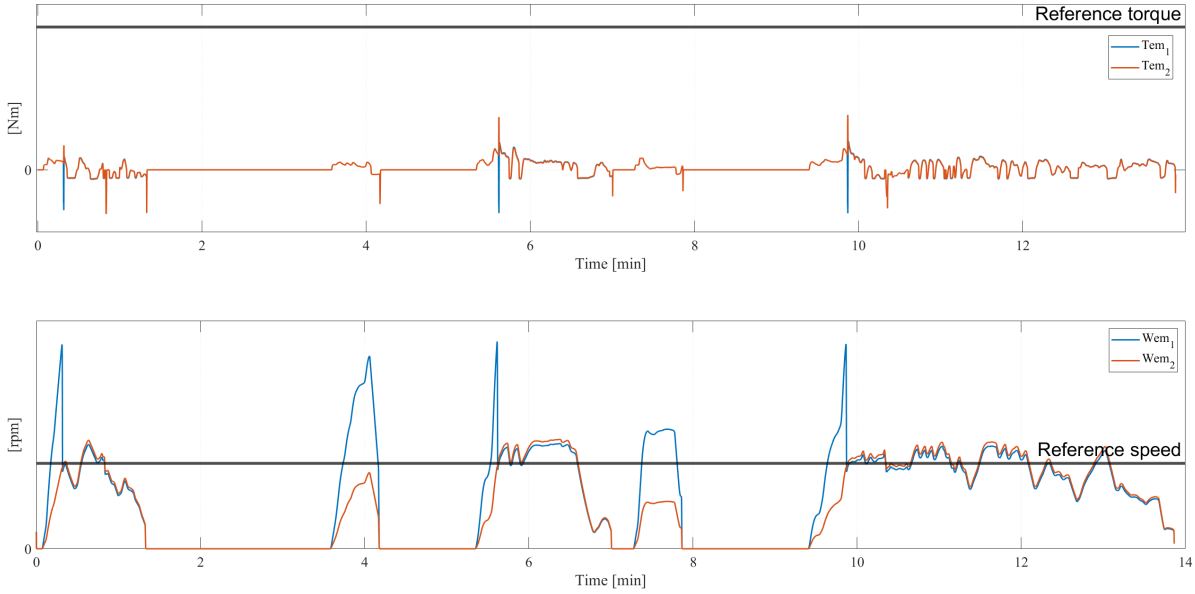


Figure 4.73: Electromagnetic torque and speed for outlying drive cycle_331, ElegantPanther_6, in the 8-ton configuration for Cluster 1.

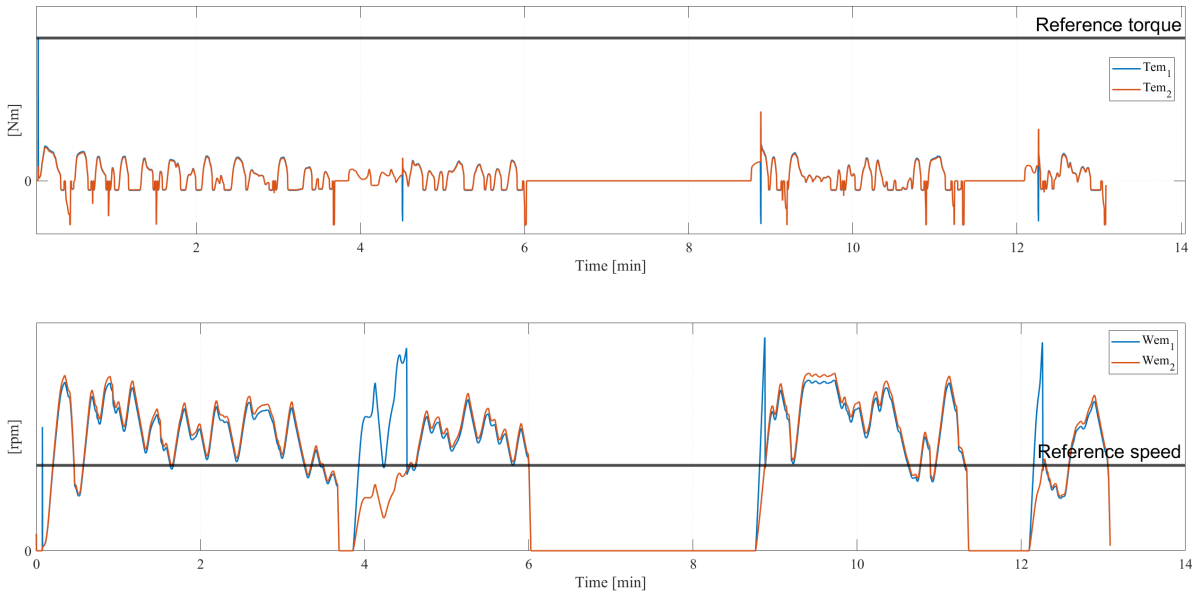


Figure 4.74: Electromagnetic torque and speed for outlying drive cycle_1207, CleverWolf_7, in the 8-ton configuration for Cluster 1.

For the 33-ton vehicle configuration, Figures 4.75- 4.76 exhibit increased torque demand. More pronounced torque fluctuations occur, as well as higher torque value peaks, reflecting the increased struggle for sustained operation associated with more mass. Another consequence of the increased mass is more frequent and larger magnitudes of regenerative events. Lastly, when comparing Figures 4.73-4.74 and Figures 4.75-4.76 both vehicle configurations exhibit balanced operation.

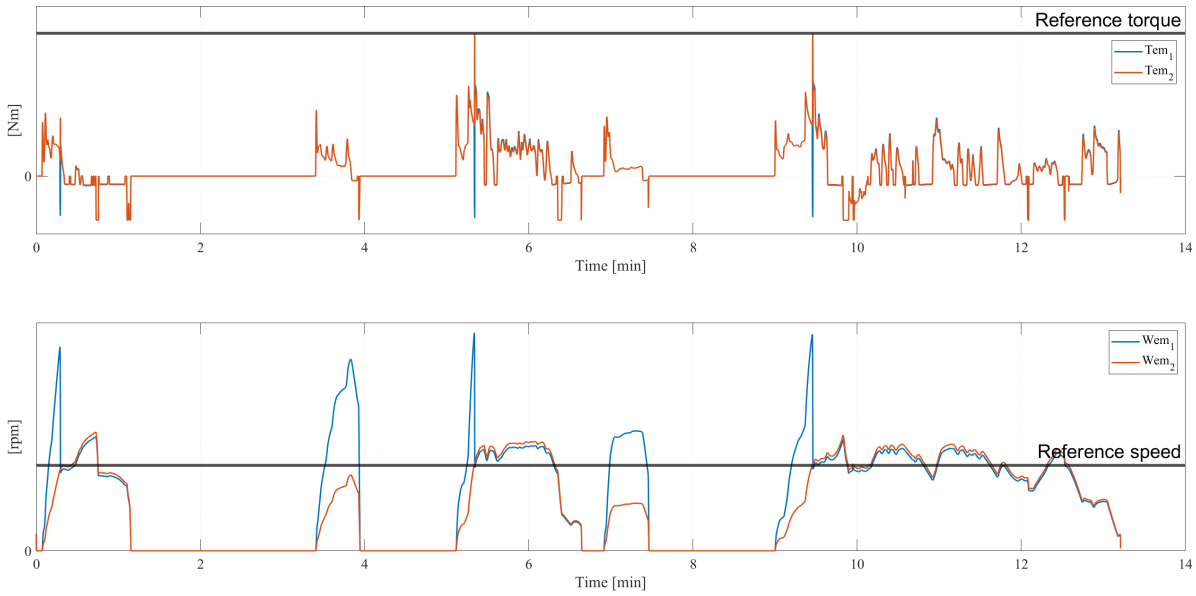


Figure 4.75: Electromagnetic torque and speed for outlying drive cycle_331, ElegantPanther_6, in the 33-ton configuration for Cluster 1.

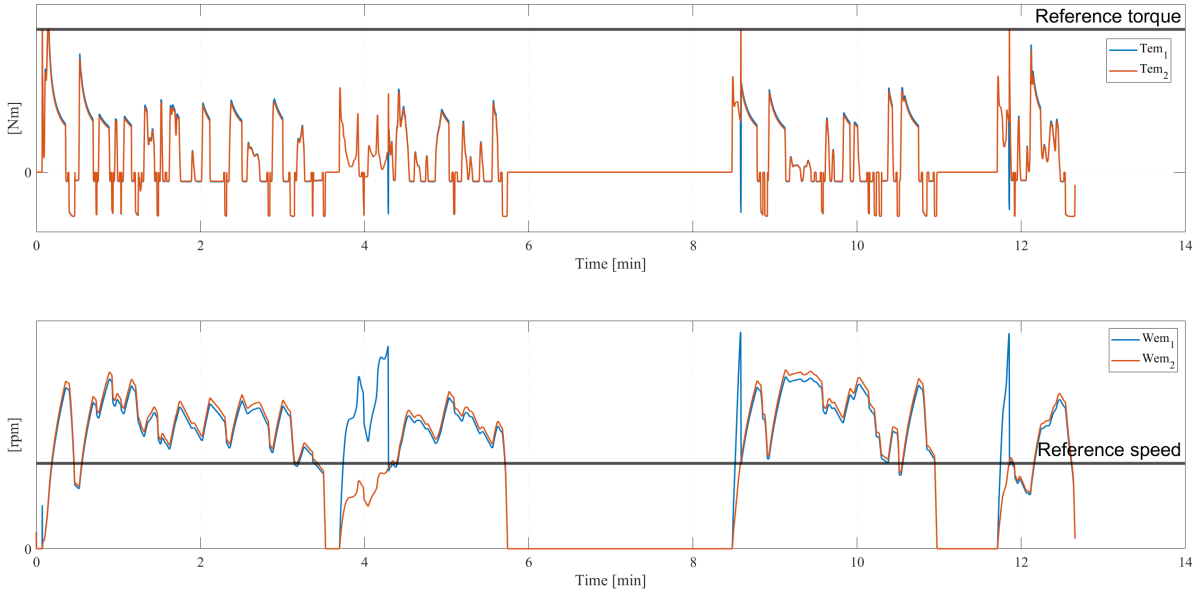


Figure 4.76: Electromagnetic torque and speed for outlying drive cycle_1207, CleverWolf_7, in the 33-ton configuration for Cluster 1.

Battery behavior

For the outlying drive cycles, in Cluster 1, Figures 4.77-4.78 highlight the configuration of 8-ton and Figures 4.79-4.80 the configuration of 33-ton. Both mass configurations exhibit repeated transient peak events for both charging and discharging. These events translate to frequent changes between power demand and regenerative braking during operation. The configuration of 33-ton exhibits higher peak-values for the C_*rate*, expected from the increased mass which also translates to larger magnitude for the regenerative braking due to increased inertia.

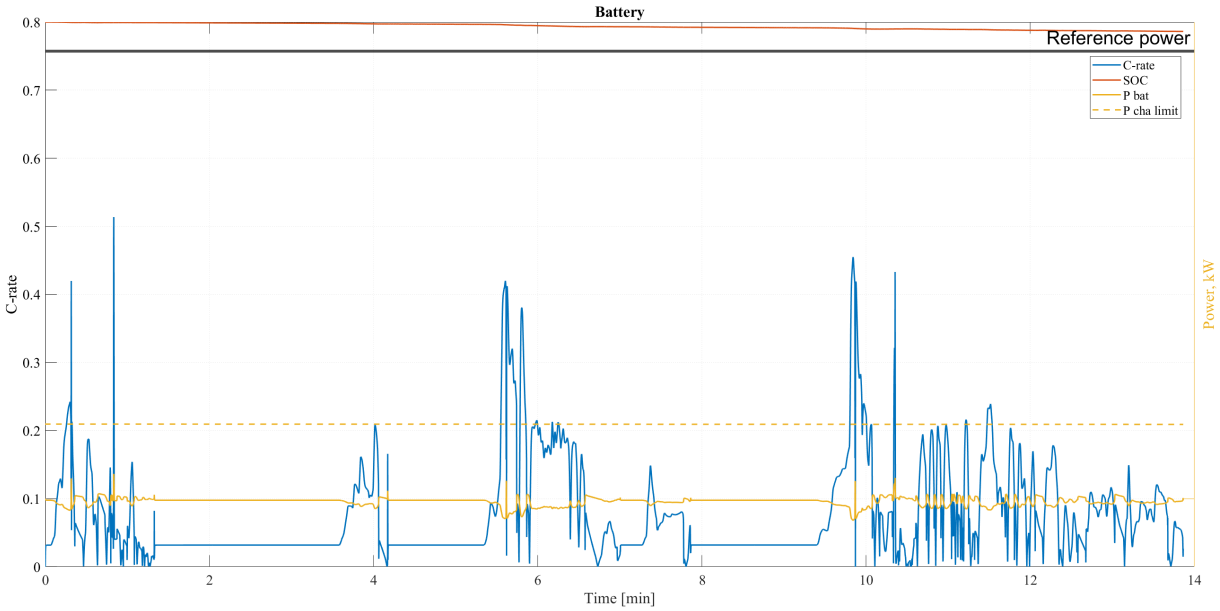


Figure 4.77: The battery behavior detected for the outlying drive cycle_331, ElegantPanther_6, in the 8-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

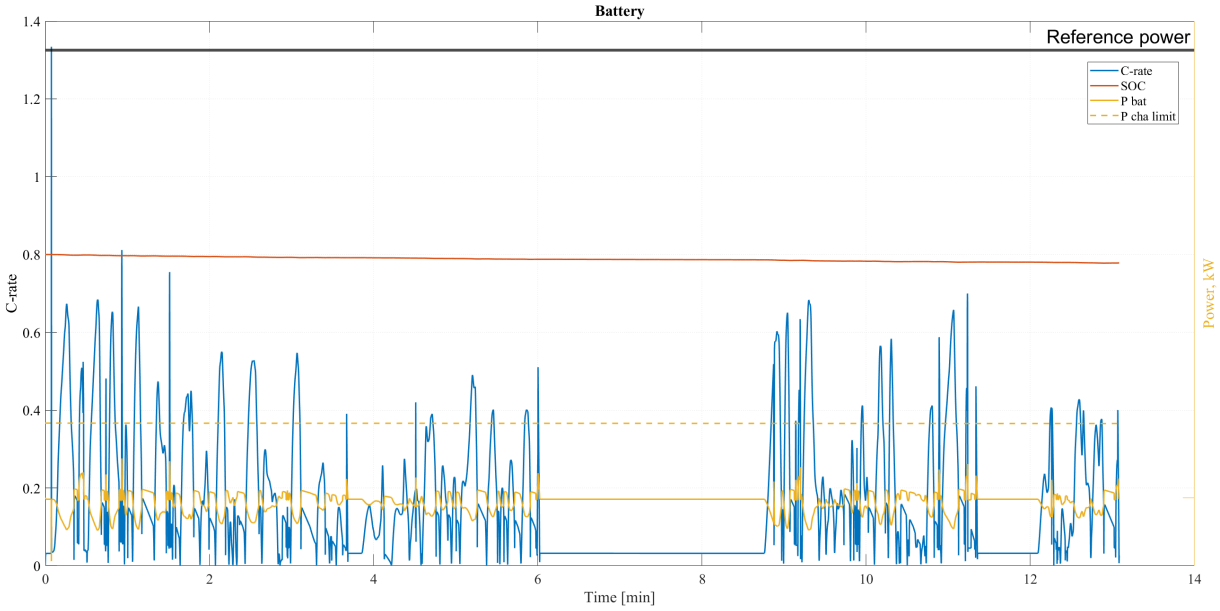


Figure 4.78: The battery behavior detected for the outlying drive cycle_1207, CleverWolf_7, in the 8-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

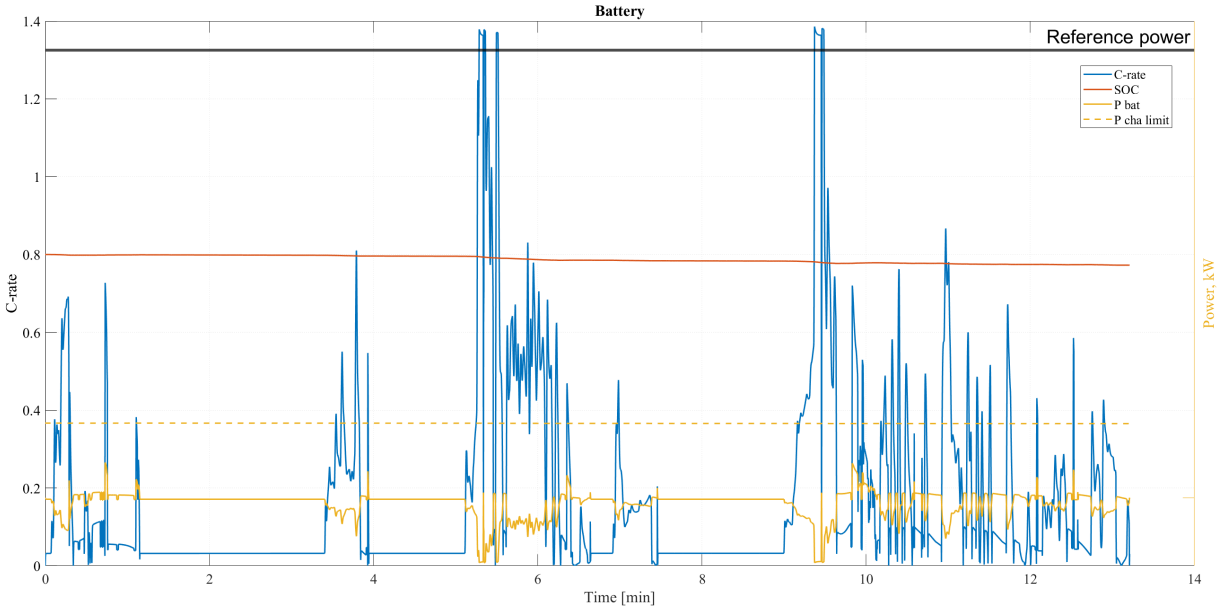


Figure 4.79: The battery behavior detected for the outlying drive cycle_331, ElegantPanther_6, in the 33-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

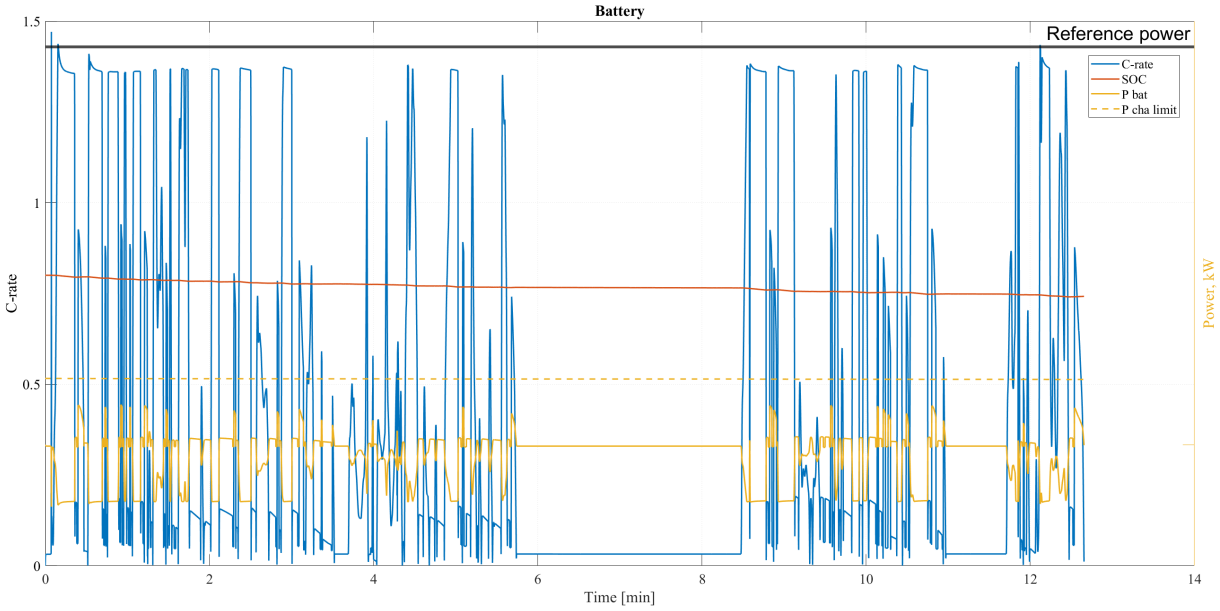


Figure 4.80: The battery behavior detected for the outlying drive cycle_1207, CleverWolf_7, in the 33-ton configuration. The yellow graph indicates how much power the battery disposes of and regenerates during each drive cycle.

Comparing the battery behavior for the outlying drive cycles with the representative ones, Figures 4.77-4.80 and 4.61-4.66, both discharge- and charging events occur more frequently and closer in time. The SOC curve remains smooth across the outlying drive cycles. Both vehicle configurations present a slow and even decrease in SOC during operation, without sudden changes. The SOC decrease is more notable for the 33-ton configuration, reflecting the increased energy consumption associated with increased mass. This also results in more energy being regained during the regenerative events, resulting in the very similar SOC curves for the different vehicle-configurations.

Comparing the two vehicle configurations for Cluster 1, a clear difference in battery stress is highlighted. The 33-ton configuration exhibits larger magnitudes- but also slightly fewer C_*rate* peaks, which is driven by the heavier power- and braking demands. The configuration of 8-ton, exhibits a wider spread of operating points but with lower magnitude values. Both configurations perform well without exhibiting continuously large magnitude C_*rate* values or extended operation near the battery power limit, indicating that even atypical drive cycles remain within the operational margin.

The detected battery behavior for the outlying drive cycles in Cluster 1, are consistent with previous results, see Section 4.3. These results state that Cluster 1 is dominated by drive cycles with high torque variance and frequent changes in acceleration and deceleration. It is also presented that Cluster 1 contains drive cycles with larger variability in usage, rather than fundamentally different drive cycles.

The energy storage data for the outlying drive cycles in Cluster 1 is presented in Table 4.40 for the 8-ton configuration and Table 4.41 for the 33-ton configuration. The tables highlight lower SOC reduction, compared to the representative drive cycles in Cluster 1. The energy throughput is increased, when compared to the representative drive cycles for Cluster 1, the ESS-efficiency remains above 98% for all drive cycles except for drive cycle_1207, for CleverWolf_7, in the 33-ton configuration.

Table 4.40: Energy storage data for the 8-ton simulations.

Parameter	**Drive cycle_331**	**Drive cycle_1207**
Integrated current [Ah]	2.83	4.46
Total Energy Output [kWh]	2.2	4.1
Total Energy captured [kWh]	0.23	0.91
ESS efficiency [%]	99.7	99.4
Energy throughput [kWh/h]	10.67	22.74
End SOC [%]	78.62	77.83
SOC difference [%-points]	1.38	2.17
Energy charge station [kWh]	0	0

Table 4.41: Energy storage data for the 33-ton simulations.

Parameter	**Drive cycle_331**	**Drive cycle_1207**
Integrated current [Ah]	5.61	11.94
Total Energy Output [kWh]	4.6	10.2
Total Energy captured [kWh]	0.70	1.99
ESS efficiency [%]	98.9	97.9
Energy throughput [kWh/h]	24.25	57.99
End SOC [%]	77.26	74.18
SOC difference [%-points]	2.74	5.82
Energy charge station [kWh]	0	0

Engine behavior

Figures 4.81-4.82 and Figures 4.83-4.84 represent the engine behavior for the 8-, respectively 33-ton configurations. Both configurations show a widely spread distribution of operating points across the torque-speed plane. In contrast to the representative drive cycles for Cluster 1, where the operating points were concentrated around moderate torque levels, the outlying drive cycles exhibit a shift towards higher torque magnitudes. This is clearest for the 33-ton configuration. The 8-ton configuration contains operating points mainly at low to moderate speeds and torques with frequent changes between traction and regenerative events. This is presented as dense patterns around the zero-torque value, indicating frequent acceleration and deceleration events. The configuration of 33-ton shifts higher into the positive torque region, reflecting the increased traction needed due to the increased mass. This is also noticeable as the negative torque magnitude reaches larger values, consistent with the increased braking force for a heavier vehicle configuration.

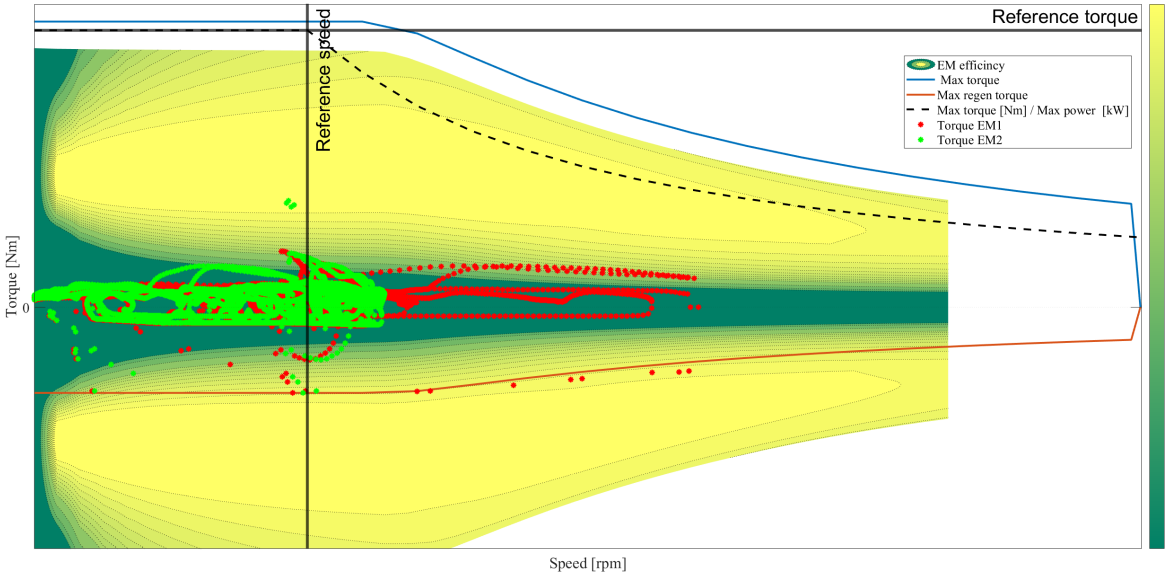


Figure 4.81: Engine behavior for the outlying drive cycle_331, ElegantPanther_6, in the 8-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

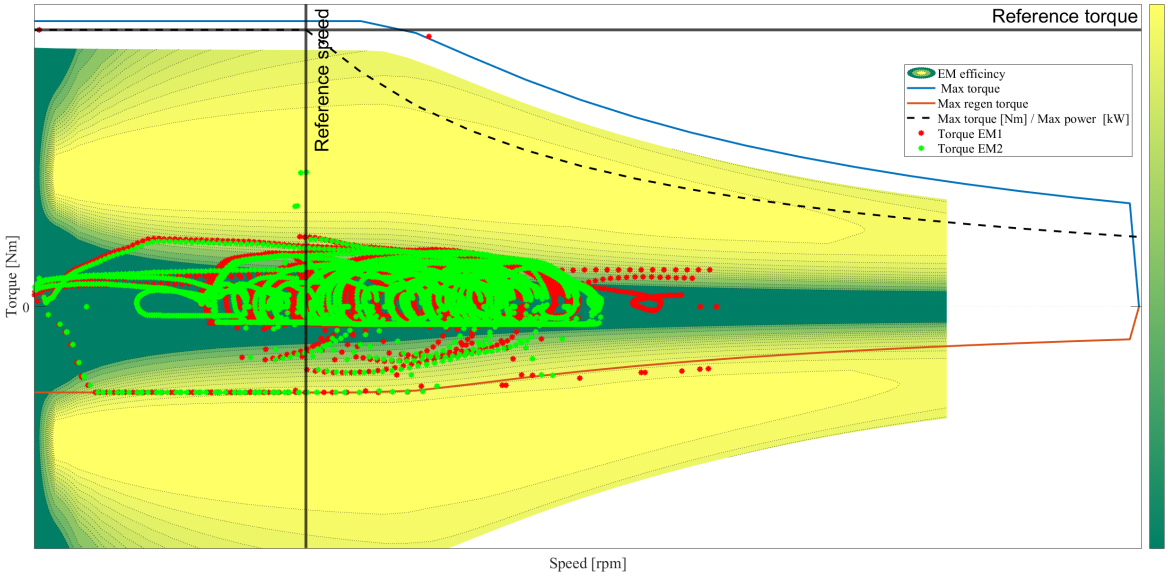


Figure 4.82: Engine behavior for the outlying drive cycle_1207, CleverWolf_7, in the 8-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

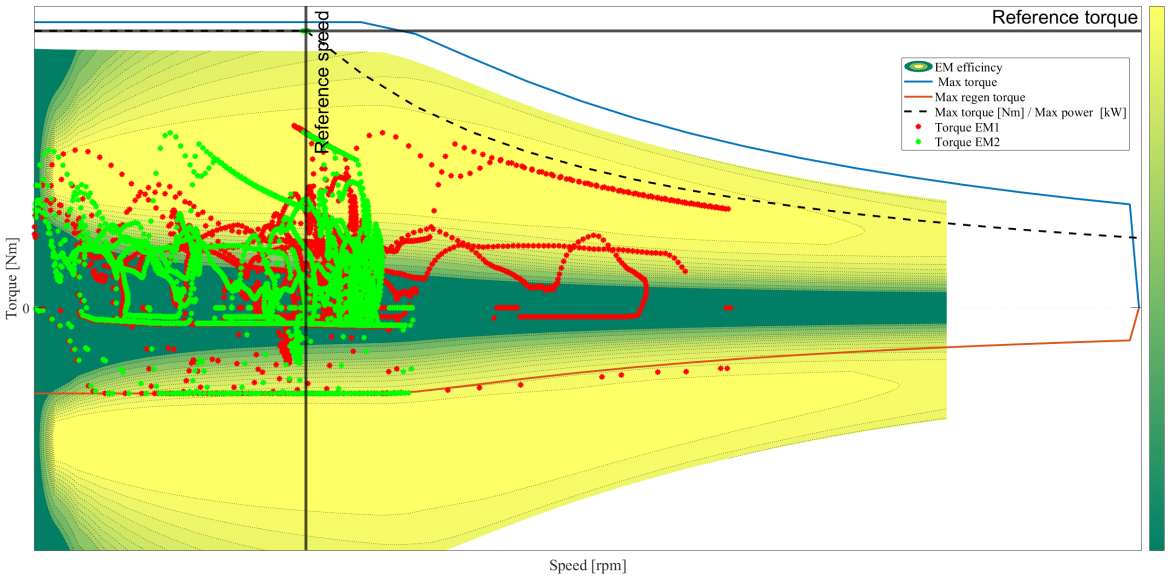


Figure 4.83: Engine behavior for the outlying drive cycle_331, ElegantPanther_6, in the 33-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

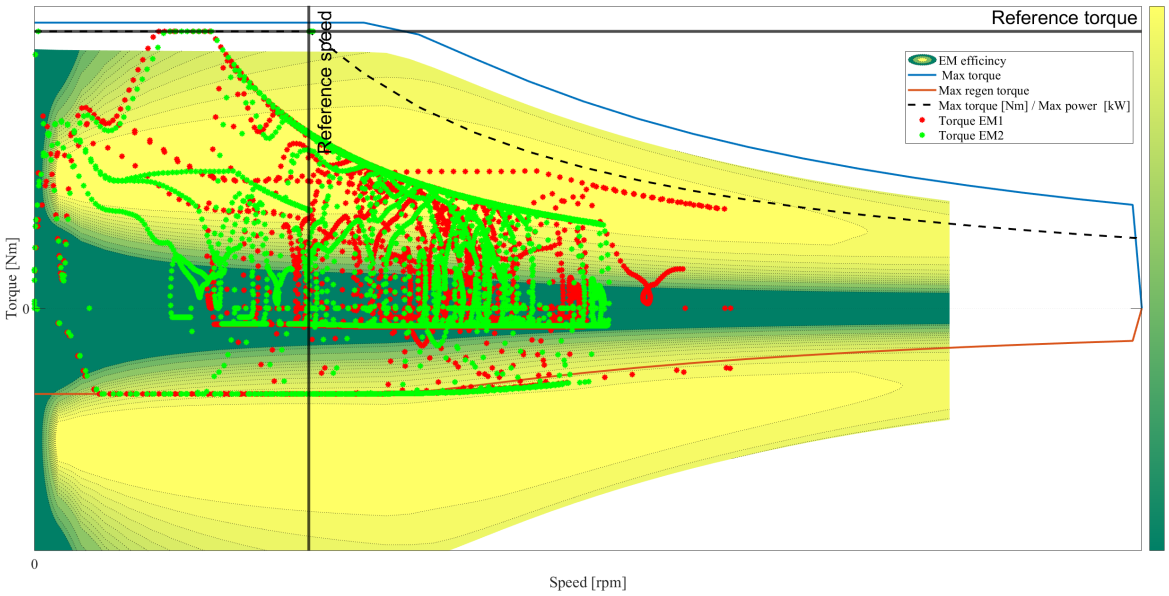


Figure 4.84: Engine behavior for the outlying drive cycle_1207, CleverWolf_7, in the 33-ton mass configuration. The colorbar indicates efficiency values from green, representing moderate levels, and yellow, representing high levels, and the dotted lines represent efficiency regions.

The majority of operating points, for both configurations, remain within the high-efficiency range of the efficiency map, typically around high efficiency, see Table 4.42 for the 8-ton configuration and Table 4.43 for the 33-ton configuration. This indicates that although the operating points are scattered wider across the torque-speed plane the engine still operates efficiently. However, when compared to the representative drive cycles the outliers present more frequent operating points towards the low efficiency regions, mostly for low speed values. These regions are clearer seen in the configuration for the 33-ton vehicle where the increased mass makes it necessary for more frequent low speed operation, due to the increased inertia. The 8-ton configuration exhibits clearer plateaus of torque for a wider range of speeds when compared to the 33-ton configuration.

Table 4.42: Electric machine efficiency data for the 8-ton simulations.

Parameter	**Drive cycle_331**	**Drive cycle_1207**
EM 1, mean efficiency [%]	85.83	88.21
EM 2, mean efficiency [%]	85.58	87.95

Table 4.43: Electric machine efficiency data for the 33-ton simulations.

Parameter	Drive cycle_331	Drive cycle_1207
EM 1, mean efficiency [%]	91.00	92.11
EM 2, mean efficiency [%]	90.64	92.00

All outlying drive cycles, regardless of mass configuration, exhibit operating points within the limits for torque and regenerative force. However, the 33-ton configuration has an increased, when compared to the representative drive cycles, amount of operating points slightly outside the limits but due to their short-lived nature, no prolonged operation occurs. These operating points are mainly due to peak power events associated with the increased mass. These extreme operating points are also present for the 8-ton configuration, but once again no continuous operation is present. The operating points confirm that even though the outliers represent more extreme and or atypical behavior, the main differences are mass related and not operational during the drive cycle.

The behavior presented in Figures 4.81-4.82 and Figures 4.83-4.84 aligns with the previous definition of Cluster 1, which is dominated by variable high torque operation, see Section 4.3. Start-stop behavior with short-lived acceleration phases and extended braking phases are dominant for the outlying drive cycles. The outlying drive cycles also highlight a shift in torque magnitude over a wider speed range without a clear change in operation. The outlying drive cycles contain increased variability in the engine torque-speed plane, for both configurations, with the operational points still remaining in high efficiency regions within the operating limits.

4.6 General discussion

The convergence towards a two cluster structure across all statistical feature based algorithms suggests that the dataset contains a dominant and stable underlying structure. This is further supported by the fact that hierarchical, centroid and medoid based methods, that rely on different optimization criteria, are all in agreement. This agreement indicates that the separation between low load and high load drive cycles is geometrically distinct in feature space. The two clusters, Cluster 0 and Cluster 1 respectively are distinguished by their difference in variance and feature ranges. Cluster 0 with its lower variance and narrower feature ranges, suggests operational consistency. Cluster 1 in contrast, exhibits higher dispersion and broader feature ranges. The broader spread of feature values suggests that demanding operations are less uniform and more context dependent. This variability is an important finding that indicates that "high load" operational behavior cannot be captured by a typical pattern and contains substructures. When the representative and outlying drive cycles of each cluster were simulated in the powertrain model, the results confirmed that the identified clusters corresponds to physically interpretable operational regimes. This further supports that the clustering interpretation as low-load and high-load, correspond to physically realistic operational behaviors.

It is however important to note that feature based clustering, based on statistical features, are inherently dependent on the selected feature representations. Since the statistical features represent a compressed and abstract version of the original time series data, the structure observed in feature space is shaped by what aspects of the original signals are preserved and those that were discarded. Consequently, the identified two-cluster structure reflects separability with respect to the selected steady-state and dynamic descriptors, rather than an absolute representation of operational behavior. Alternative feature representations could potentially emphasize different aspects of variability and reveal additional structures.

The consistent identification of two dominant operating regimes indicates that feature based clustering is sufficient for extracting the primary operational structure of the dataset. Given its computational efficiency, interpretability, and direct connection to physically meaningful features, it provides a practical framework for identification of drive cycles. In contrast, the CAE reveals finer temporal submodes, at the cost of higher model complexity and reduced interpretability. The suitability of each method therefore depends on the objective: feature based clustering is appropriate for dominant regime categorization and simulation linkage, while approaches such as the CAE are better suited for detailed temporal variability analysis.

Although explicit optimization of component sizing and/or control strategies was outside the scope of this thesis, the combined clustering and simulation framework provides insight into how the powertrain is utilized under the investigated operational regimes. This structured understanding of usage patterns forms a necessary foundation for future optimization studies by identifying the operating conditions that govern component stress and performance. Analysis across the 8- and 33-ton

configurations, representing the extreme application weights within operational limits, shows that the qualitative behavior of a given drive cycle remains consistent regardless of vehicle mass. Although a few operating points occur at limit levels, no sustained operation at these limits is observed in either of the two simulated mass configurations of the powertrain. This suggests that, under the investigated conditions, the powertrain is not operated near its predefined limits for extended periods. Deviations in *Travel time*, observed during simulations, were primarily due to interpolation and numerical rounding effects. This indicates that mass primarily affects the magnitude of the system response rather than the operational regime itself. Consequently, cluster separation is not driven by sustained operation at engine or battery limits, but by differences in variability and load distribution. Driving behavior and transient complexity therefore appear to be the dominant factors defining cluster identity and operational stress.

The simulation results also demonstrate that analysis of a single component is insufficient to fully characterize operational behavior. For example, only examining the engine efficiency map for certain drive cycles, such as drive cycles 886 and 196 in the 8-ton configuration, Figures 4.33 and 4.49, the distribution of operating points appears very similar. Both drive cycles exhibit operating points concentrated in the same region of the torque-speed plane, forming a comparable elongated distribution. The relative positioning of operating points, for engine one and engine two, is very similar across the two drive cycles. The main difference is quantitative rather than qualitative, drive cycle_886 contains substantially more operating points than drive cycle_196, but the underlying shape and location of the distribution within the efficiency map is preserved. This geometric similarity means that, when examining the efficiency map alone, drive cycle_196 appears to be a scaled down version of drive cycle_886, as if the same operational pattern occurred, but fewer times. However, when engine behavior, Figures 4.21 and 4.37, and battery behavior, Figures 4.27 and 4.43, are included, differences in powertrain behavior become evident. Figures 4.21 and 4.37, highlight very different engine speed magnitudes where drive cycle_886 operates at higher speeds for longer periods than drive cycle_196. Figures 4.27 and 4.43 exhibits different C_*rate*, where drive cycle_886 uses more current and for longer periods than drive cycle_196, during the drive cycles. This emphasizes the need to analyze components together when interpreting drive cycle characteristics and supports the methodological choice of using multidimensional feature extraction and clustering.

Finally, it is important to note that the clusters do not represent two perfectly separated categories. Some drive cycles exhibit characteristics that are partially shared between clusters. This transitional behavior reflects the continuous nature of real-world driving, strengthening the real world applicability of the framework.

5
Conclusion

This thesis has presented a data driven approach for the identification, clustering, and characterization of heavy duty vehicle drive cycles using unsupervised learning techniques. By combining signal preprocessing, feature extraction, multiple clustering algorithms, and simulations, the proposed framework enables a structured analysis of real world drive cycle variability.

HAC, K-Means, K-Medoids, and a CAE were evaluated using internal validation metrics and cluster agreement measures. While the classical clustering methods consistently captured a dominant two cluster structure in the time series data, the CAE revealed additional operational sub patterns with higher internal variability. K-Means and K-Medoids were found to provide robust and interpretable selections of representative drive cycles. The strong agreement between these clustering approaches further supports the reliability of the identified structure, consisting of a low-variance, steady operational regime and a high-variance, load-intensive regime characterized by broader torque and current distributions.

The clustering analysis revealed distinct operational driving patterns, reflected in differences in speed profiles, torque demand, and electrical current characteristics. The identification of representative and outlying drive cycles for each cluster enabled a focused investigation of both typical and extreme operating conditions. Simulation based evaluation of these drive cycles highlighted differences in engine behavior, such as engine torque- and speed demand, battery dynamics across clusters and vehicle mass configurations.

The simulated drive cycles exhibited behaviors consistent with their respective clusters. The simulations reveal a powertrain that is designed with large operational margins, as only a limited number of operating points occur at the defined limit levels for the analyzed drive cycles. This suggests potential opportunities for better utilization of the application and/or space for resizing, although explicit redesign was not evaluated within the scope of this thesis.

The results demonstrate that representative drive cycle selection based on clustering can serve as a valuable tool for structured powertrain evaluation. By separating operating regimes, such as stable and high-variability regimes, and linking them to simulation outcomes, the framework provides a systematic approach of component utilization, stress exposure, and operational margins. This structured characterization of the powertrain provides a foundation for future optimization studies, includ-

ing component sizing, control strategy refinement, and scenario-based validation of electrified powertrains in heavy-duty vehicles.

6
Future Work

Several directions for future work emerge from the findings and limitations of this thesis. One natural extension would be to further investigate unsupervised drive cycle design, by for example reducing the reliance on manually selected parameters. Utilizing a systematic exploration of pre-processing thresholds and filtering, such as automated threshold selection or stability based grid searches, could improve the robustness and reproducibility of the pipeline. Such refinements may improve consistency across other datasets, engines, and operational contexts.

Another important aspect for future research is the evaluation of the proposed pipeline across multiple applications. In this thesis, the methodology has been applied to a single application. Applying the same pipeline to different vehicles, engine configurations or duty cycles would provide insight into its ability to generalize and its limitations. This would help determine whether the observed clustering structure reflects application specific characteristics or more general operational patterns.

Further work could also include a broader comparison of clustering algorithms with fundamentally different assumptions and architectures. While this thesis focuses on distance based and partitioning approaches, alternative methods such as density based clustering or probabilistic models may reveal other structures in the data. Investigating such methods could prove useful by finding additional perspectives on operational regimes.

The CAE represents another promising direction for continued research. In this work, it was primarily used as an exploratory tool to find deeper temporal patterns in the raw time series. These findings indicate that there is value in more features and signals passed through the feature based clustering algorithms, to capture the true complexity of the time series. Future studies could investigate deeper architectures, larger latent spaces, and training strategies that rely more directly on the raw signal data. With further refinement, autoencoder based representations may enable more accurate separation and identification of subtle operating regimes and transitions.

Another aspect to consider for future research could be to use a set of predetermined drive cycles, for example the representative ones of a cluster, and investigate the effect of different powertrain components. By doing so, a possibility arises for an optimized powertrain utilized for a specific type of operation could be determined. Which could prove favorable for both maintenance and application development.

Finally, it is important to acknowledge that the clustering and simulation results are influenced by decisions made in the early stages of the methodology, including signal pre-processing, feature selection, normalization, and dimensionality reduction. These choices introduce subjectivity and may bias the resulting clusters toward certain interpretations of the data. Future work could therefore focus on quantifying the sensitivity of the clustering outcomes to these design choices, for example through systematic studies or alternative pre-processing pipelines.

Bibliography

[1] X. He, D. Madigan, B. Yu, and J. Wellner, *Statistics at a crossroads; who is for the challenge?* 2025. arXiv: 2503.22945 [stat.OT]. [Online]. Available: https://arxiv.org/abs/2503.22945.

[2] M. Alizadeh and J. Ma, "High-dimensional time series analysis and anomaly detection: A case study of vehicle behavior modeling and unhealthy state detection," *Advanced Engineering Informatics*, vol. 57, p. 102 041, 2023, ISSN: 1474-0346. DOI: 10.1016/j.aei.2023.102041. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474034623001696.

[3] European Commission, Directorate-General for Climate Action. "Going climate-neutral by 2050: A strategic long-term vision for a prosperous, modern, competitive and climate-neutral eu economy." Publications Office of the European Union, ISBN 978-92-76-02037-0, Catalogue number ML-04-19-339-EN-N. (2019), [Online]. Available: https://op.europa.eu/en/publication%E2%80%91detail/-/publication/92f6d5bc-76bc-11e9-9f05-01aa75ed71a1/language-en.

[4] R. Hellrand and J. Malmer Göransson, "Defining, analyzing, and clustering drive cycles for engine applications through feature engineering," Accessed: 2025-11-17, Master's thesis, Chalmers University of Technology, 2025. [Online]. Available: https://odr.chalmers.se/items/1ac4af95-5923-46e2-9471-2a9766658033.

[5] T. W. Anderson, *The Statistical Analysis of Time Series*. New York / London: John Wiley Sons, 1971, Wiley Classic Library edition, ISBN: 0-471-02900-9.

[6] National Institute of Standards and Technology (NIST) / SEMATECH, *Univariate time series models (section 6.4.4)*, Accessed: 2025-11-17, NIST, 2003, ch. 6.4.4.

[7] E. Inc., *Multivariate time series - an overview*, https://www.sciencedirect.com/topics/computer-science/multivariate-time-series, Accessed: 2025-11-17, ScienceDirect / Elsevier, n.d.

[8] A. Inc., What is an EV Powertrain? https://www.ansys.com/simulation-topics/what-is-an-ev-powertrain, Accessed: 2025-11-17, ANSYS Inc., 2024.

[9] J. G. Hayes and G. A. Goodarzi, *Electric Powertrain: Energy Systems, Power Electronics and Drives for Hybrid, Electric and Fuel Cell Vehicles*. Hoboken, NJ, USA: John Wiley & Sons, 2018, p. 560, ISBN: 9781119063643.

[10] Volvo Penta. "Electromobility solutions for off-highway industrial applications." Accessed: 2026-02-10, Volvo Penta. (2026), [Online]. Available: https://www.volvopenta.com/industrial/industrial-engines/electromobility/.

[11] Z. Rahman, M. Ehsani, and K. L. Butler, "An investigation of electric motor drive characteristics for ev and hev propulsion systems," *SAE Transactions*, vol. 109, no. 6, pp. 2396–2403, 2000. [Online]. Available: https://www.jstor.org/stable/44687077.

[12] E. Roshandel, A. Mahmoudi, S. Kahourzade, and W. L. Soong, "Efficiency maps of electrical machines: A tutorial review," *IEEE Transactions on Industry Applications*, vol. 59, no. 2, pp. 1263–1272, 2023. DOI: 10.1109/TIA.2022.3210077. [Online]. Available: https://ieeexplore.ieee.org/document/9904865.

[13] C. Su, H. Kim, J. Lee, H. Chung, and S.-K. Sul, "Drive control of an electric vehicle based on road slope estimation using a fuzzy controller," in *2015 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*, IEEE, 2015, pp. 593–598. DOI: 10.1109/SPEEDAM.2015.7310051. [Online]. Available: https://ieeexplore.ieee.org/document/7310051.

[14] Y. Zhang, H. Liu, J. Wang, and W. Zhao, "Thermal modeling and analysis of traction motors for electric vehicles," *IEEE Transactions on Transportation Electrification*, vol. 6, no. 4, pp. 1541–1552, 2020. DOI: 10.1109/TTE.2020.3030469. [Online]. Available: https://ieeexplore.ieee.org/document/9247471.

[15] M. Bauer, M. Hoffmann, and M. Lienkamp, "Energy efficiency assessment of electric powertrains under real driving conditions," *Energy Technology*, 2024. DOI: 10.1002/ente.202401331. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/ente.202401331.

[16] SAE International, *Sae mobilus® — what we offer*, https://www.saemobilus.org/what-we-offer, Accessed: 2025-11-17, SAE International, n.d.

[17] M. K. Kondaru, K. P. Telikepalli, S. V. Thimmalapura, and N. K. Pandey, "Generating a real world drive cycle–a statistical approach," in *WCX World Congress Experience 2018*, ser. SAE Technical Papers, Warrendale, PA, USA: SAE International, 2018. DOI: 10.4271/2018-01-0325. [Online]. Available: https://www.sae.org/papers/generating-a-real-world-drive-cycle-a-statistical-approach-2018-01-0325.

[18] T. Verdonck, B. Baesens, M. Óskarsdóttir, and S. vanden Broucke, "Special issue on feature engineering – editorial," *Machine Learning*, vol. 113, pp. 3917–3928, 2021. DOI: 10.1007/s10994-021-06042-2. [Online]. Available: https://doi.org/10.1007/s10994-021-06042-2.

[19] C. C. Aggarwal, *Clustering: The Definitive Guide*. Cham, Switzerland: Springer Nature Switzerland, 2019, Accessed: 2025-11-17, ISBN: 978-3-030-20391-4. [Online]. Available: https://books.google.se/books?hl=en&lr=&id=kYC3YCyl_tkC&oi=fnd&pg=PR5&dq=clustering.

[20] O. Maimon and L. Rokach, Eds., *Data Mining and Knowledge Discovery Handbook*, 1st ed. New York, NY, USA: Springer New York, 2005, ISBN: 978-0-387-25465-4. DOI: 10.1007/b107408. [Online]. Available: https://doi.org/10.1007/b107408.

[21] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philosophical Transactions of the Royal Society A: Math-*

ematical, Physical and Engineering Sciences, vol. 374, no. 2065, p. 20 150 202, 2016. DOI: 10.1098/rsta.2015.0202. [Online]. Available: https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202.

[22] T. Breitenbach, L. Rasbach, C. Liang, and P. Jahnke, "A principal feature analysis," *Journal of Computational Science*, vol. 58, p. 101 502, 2022, ISSN: 1877-7503. DOI: https://doi.org/10.1016/j.jocs.2021.101502. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877750321001666.

[23] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, p. 1295, 2020, ISSN: 2079-9292. DOI: 10.3390/electronics9081295. [Online]. Available: https://www.mdpi.com/2079-9292/9/8/1295.

[24] GeeksforGeeks, *Euclidean Distance – GeeksforGeeks*, https://www.geeksforgeeks.org/maths/euclidean-distance/, Accessed: 2026-03-03, 2026.

[25] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. MIT Press, 2009, [Skillsoft version]. [Online]. Available: https://chalmers.skillport.com/skillportfe/main.action?assetid=49924 (visited on 03/09/2026).

[26] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proceedings of the 18th Annual ACM–SIAM Symposium on Discrete Algorithms (SODA 2007)*, 2007, pp. 1027–1035. DOI: 10.1145/1283383.1283494. [Online]. Available: https://theory.stanford.edu/~sergei/papers/kMeansPP-soda.pdf.

[27] X. Jin and J. Han, "K-medoids clustering," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2010, pp. 564–565, ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_426. [Online]. Available: https://doi.org/10.1007/978-0-387-30164-8_426.

[28] R. P. Adams, "Hierarchical clustering (cos 324 – elements of machine learning, princeton university)," Princeton University, Course Notes / Lecture Handout, 2018, Accessed: 2025-11-17. [Online]. Available: https://www.cs.princeton.edu/courses/archive/fall18/cos324/files/hierarchical-clustering.pdf.

[29] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," *arXiv preprint*, vol. arXiv:1109.2378, 2011, Accessed: 2025-11-17. [Online]. Available: https://arxiv.org/pdf/1109.2378.

[30] scikit-learn developers, *Sklearn.cluster.agglomerativeclustering*, version 1.4.2, Accessed: 2025-11-17, scikit-learn, 2024. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html.

[31] J. Noble, *What is hierarchical clustering?* https://www.ibm.com/think/topics/hierarchical-clustering, Accessed: 2025-11-17, IBM, 2023.

[32] I. D. Mienye and T. G. Swart, "Deep autoencoder neural networks: A comprehensive review and new perspectives," *Archives of Computational Methods in Engineering*, vol. 32, pp. 3981–4000, 2025. DOI: 10.1007/s11831-025-10260-5. [Online]. Available: https://link.springer.com/article/10.1007/s11831-025-10260-5.

[33] "Convolutional auto-encoder." Accessed: 2026-02-09, ScienceDirect. (2022), [Online]. Available: https://www.sciencedirect.com/topics/computer-science/convolutional-auto-encoder.

[34] A. Alqahtani, M. Ali, X. Xie, and M. W. Jones, "Deep time-series clustering: A review," *Electronics*, vol. 10, no. 23, p. 3001, 2021. DOI: 10.3390/electronics10233001. [Online]. Available: https://www.mdpi.com/2079-9292/10/23/3001.

[35] scikit-learn developers, *Clustering — calinski–harabasz index*, https://scikit-learn.org/stable/modules/clustering.html#calinski-harabasz-index, Accessed: 2025-11-17, scikit-learn, 2025.

[36] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974. DOI: 10.1080/03610927408827101. eprint: https://www.tandfonline.com/doi/pdf/10.1080/03610927408827101. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/03610927408827101.

[37] scikit-learn developers, *Sklearn.metrics.davies_bouldin_score*, https://scikit-learn.org/stable/modules/generated/sklearn.metrics.davies_bouldin_score.html, Accessed: 2025-11-17, scikit-learn, 2025.

[38] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *Journal of Intelligent Information Systems*, vol. 17, Oct. 2001. DOI: 10.1023/A:1012801612483.

[39] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999. DOI: 10.1145/331499.331504.

A
Appendix 1

This appendix contains the results from the K-Medoids clustering.

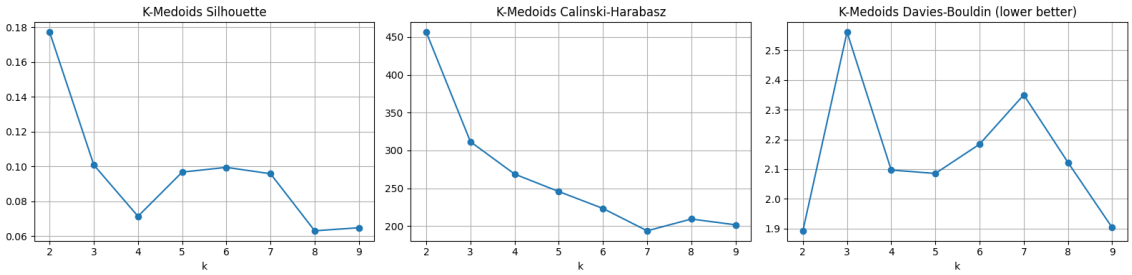


Figure A.1: Evalutaion metrics for the K-Medoids clustering.

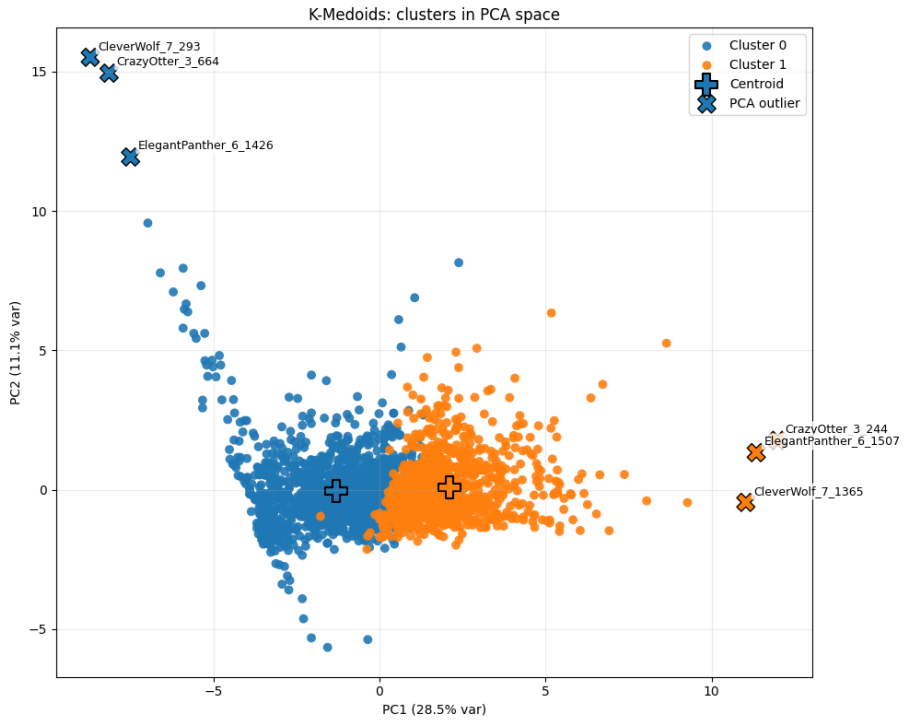


Figure A.2: K-Medoids clusters in PCA space.

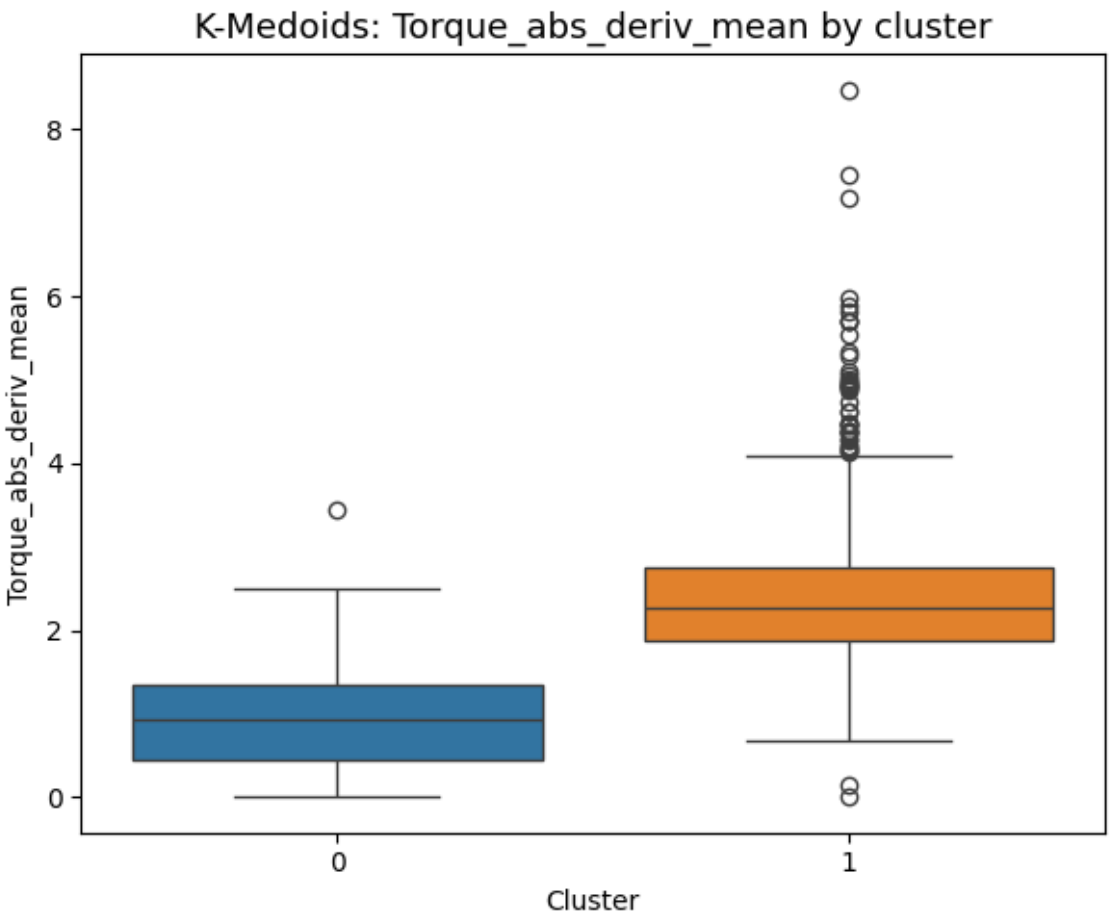


Figure A.3: Boxplot of the Torque|Δ| for the K-Medoids algorithm.

B

Appendix 2

This appendix provides additional results excluded from the Results section 4.

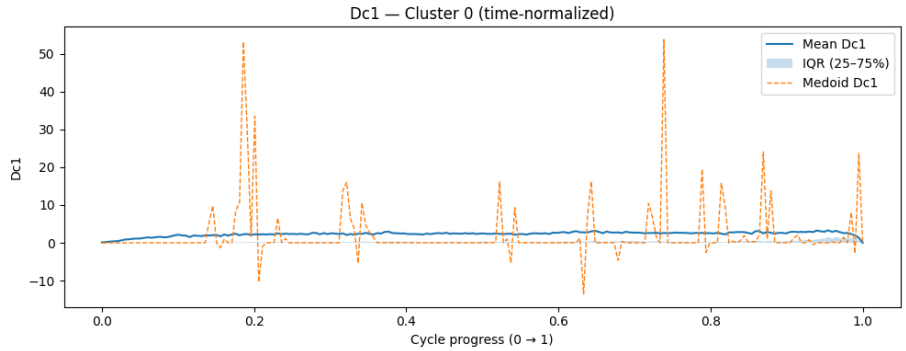


Figure B.1: Time series of the most representative cycle for cluster 0, with mean, IQR and Medoid for DC value.

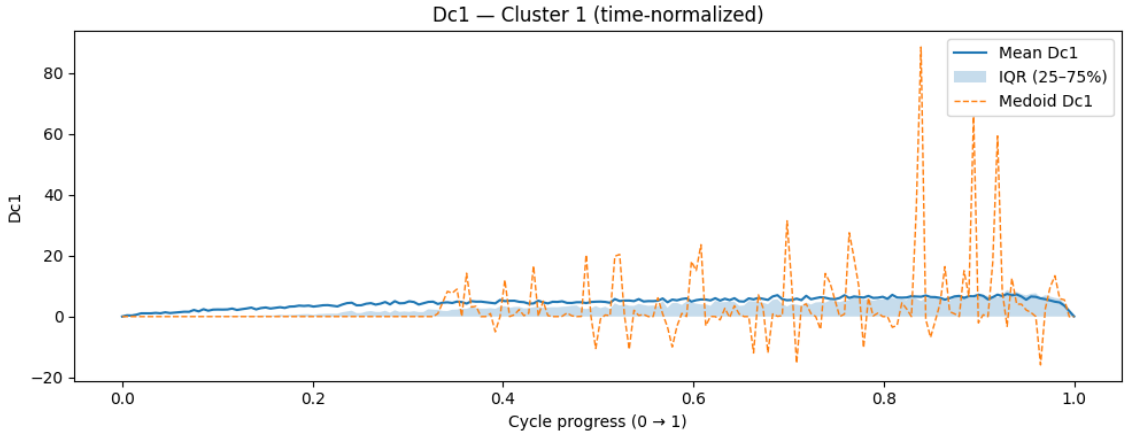


Figure B.2: Time series of the most representative cycle for cluster 1, with mean, IQR and Medoid for DC value.

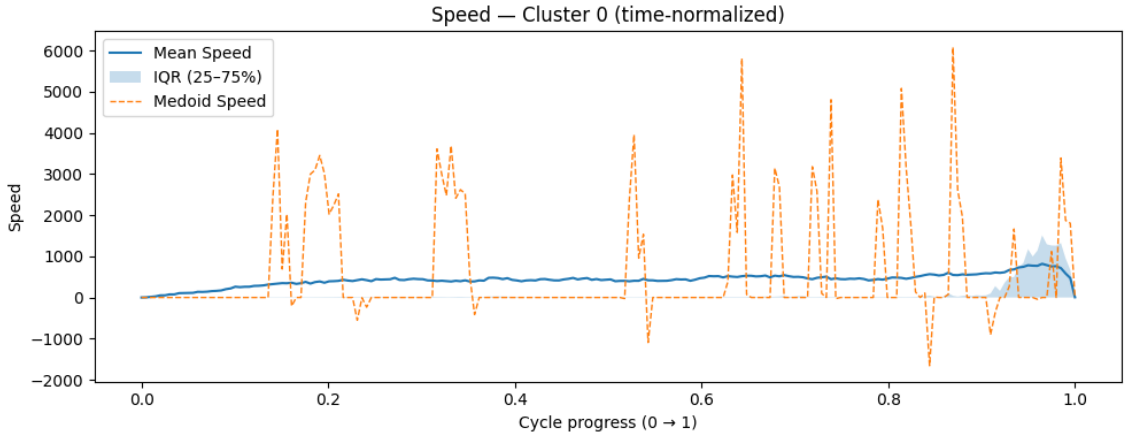


Figure B.3: Time series of the most representative cycle for cluster 0, with mean, IQR and Medoid for speed value.

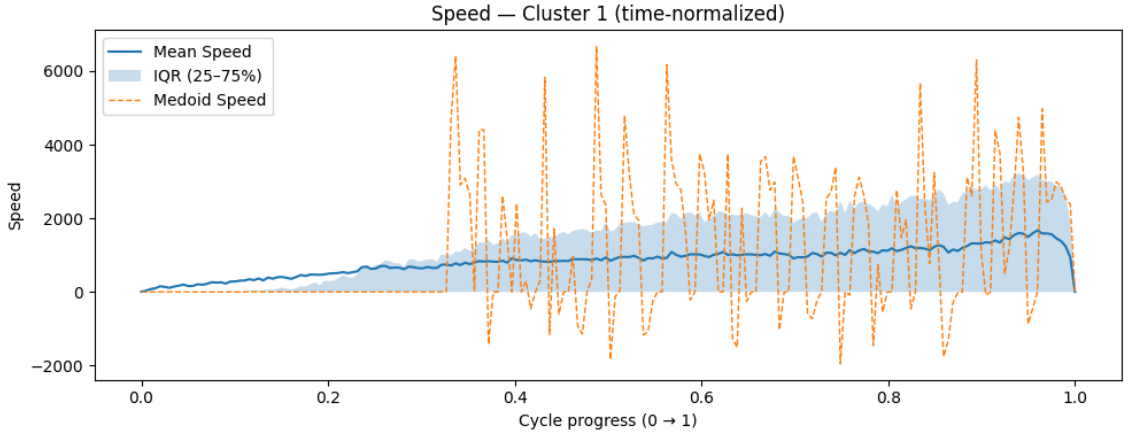


Figure B.4: Time series of the most representative cycle for cluster 1, with mean, IQR and Medoid for speed value.

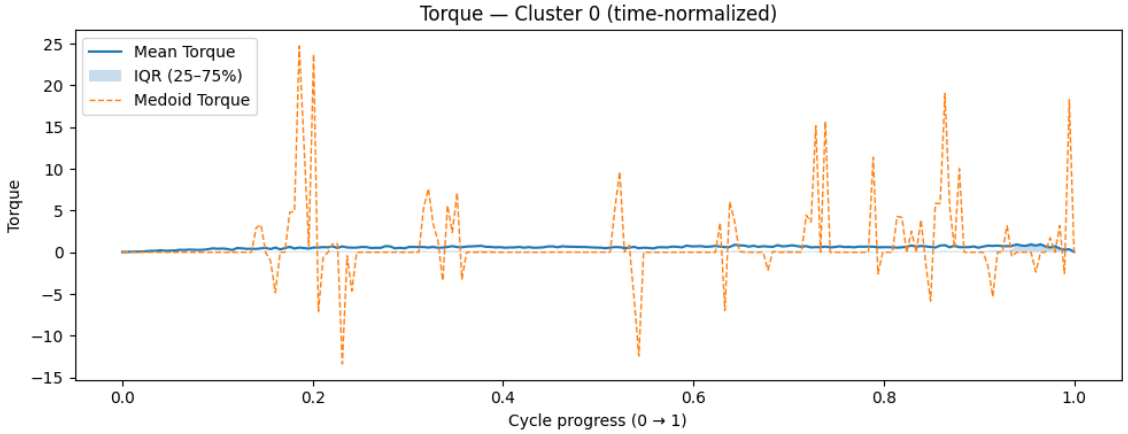


Figure B.5: Time series of the most representative cycle for cluster 0, with mean, IQR and Medoid for torque value.

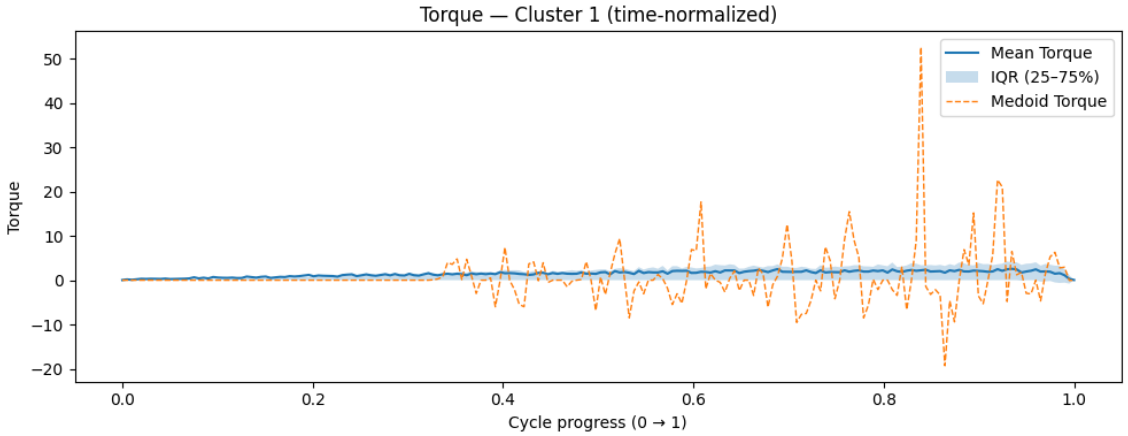


Figure B.6: Time series of the most representative cycle for cluster 1, with mean, IQR and Medoid for torque value.

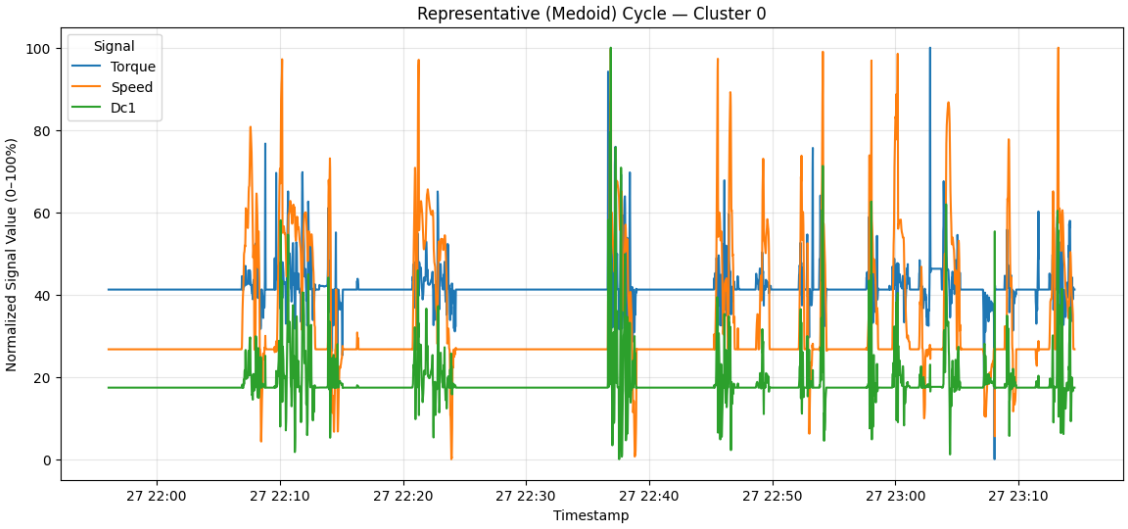


Figure B.7: Time series for the three signals: Torque, speed and DC1 of the most representative cycle in Cluster 0.

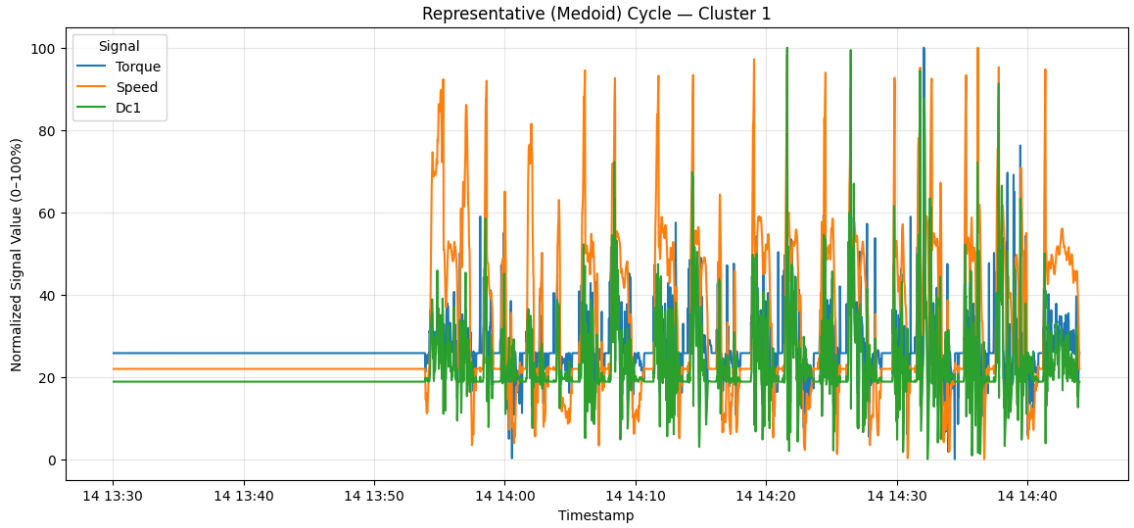


Figure B.8: Time series for the three signals: Torque, speed and DC1 of the most representative cycle in Cluster 0.

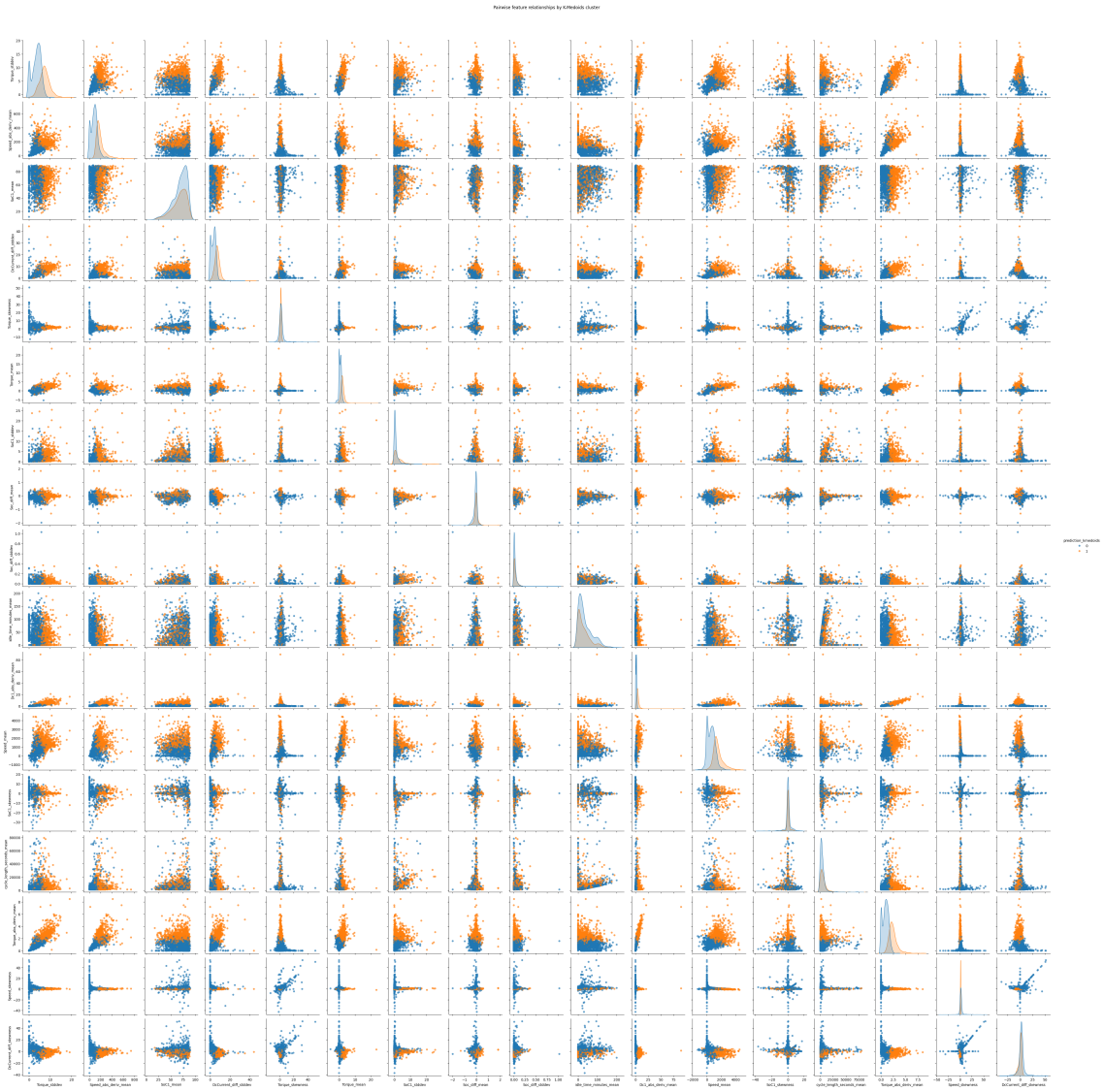


Figure B.9: Complete pairplot of the features used in the clustering analysis for the K-Medoids algorithm.

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY